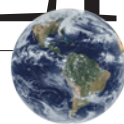


THIS WEEK

EDITORIALS

SEXISM An update on Nature's efforts to include women **p.188**

WORLD VIEW Charisma of the European Research Council **p.189**



STORMS Satellite data could help predict hurricane strength **p.191**

Failure of care

Laboratory animals must have the very best standard of care if we are to justify their use in science. As one institution is found wanting, others should look to review their animal-welfare practices.

In 1911, the United Kingdom made it a criminal offence for scientists, doctors and others to 'infuriate' animals in their care. It was a curious choice of words, and one that demonstrates the level of protection that British authorities sought — and still seek — to give rats, mice and other creatures. Such animals are vulnerable, and scientists who use them in laboratories, British law demands, must work to protect not just their physical well-being, but their emotional state too. As a result, researchers and those who support their efforts, including this journal, can respond to critics of the ethics of vivisection with the twin defences that the work is essential and carried out under strict controls.

It is infuriating, then, to see that legitimate defence of the use of animals in research weakened, as it has been this week by revelations of poor animal-research standards at one of the world's leading universities. The university, a new report finds, lacks adequate leadership, management, operational, training, supervisory and ethical-review systems to support high standards in animal use and welfare. It highlights an academic culture of complacency, of understaffed animal facilities, patchy training and poor supervision. It makes difficult reading for anyone who has ever argued for the use of animals in research.

The institution criticized is Imperial College London, but the failings and the problem almost certainly extend further. At a press conference held to release the report in London on 9 December, members of the investigating committee were keen to emphasize that they had witnessed no malpractice themselves during their inspections, but said that the improvements they suggested would make malpractice less likely to occur in future, at Imperial College or elsewhere. To manage and coordinate hundreds of scientists and support staff involved in work with animals is a complex task — one, the report says, that universities should take as seriously as running an individual academic department.

To be clear, the failings identified do not weaken the case for using animals in science — although the more strident opponents of vivisection will probably argue otherwise — but instead offer a clear demonstration that the ethical considerations of such work do not begin and end with its scientific justification. The day-to-day implementation of the task is as important as the purity of the goal. Those who take the moral high ground must, after all, be sure they stand on solid foundations.

The value of research with animals and the need for that research to be done properly should not be confused — and it is unfortunate that Imperial College chose to buttress its contrite official response to the review with polished summaries of the benefits of animal research for human health. That is not what this is about.

Imperial College has already said that it accepts all the recommendations of the committee that wrote the critical report, and is working to improve the situation. The failings identified are unlikely to be picked up in government inspections, and other institutions should carry out their own reviews. Imperial College commissioned the current report

only after specific allegations of misconduct were raised by anti-vivisection campaigners. Other universities should not need to be asked twice.

In the case of Imperial College, the investigating committee had to tip-toe around the specific allegations, raised in April after an undercover investigation by the British Union for the Abolition of Vivisection. Those allegations are being looked at in a separate inquiry by the British government, which grants scientists and universities the licences necessary to work with animals, and sets out the required conditions.

"The ethical considerations of animal research do not begin and end with its scientific justification."

Not for the first time when it comes to academia, many of the proposed improvements centre on better and more constructive communication. E-mailed queries to project scientists from the technicians who supervise the animals, and who are best placed to spot if they are in discomfort, should be replaced with face-to-face meetings. Statisticians should be brought in to discuss projects at the planning stage, so as to ensure that the experiments will be properly powered and the results meaningful.

What is most disappointing about the weaknesses exposed at Imperial College is that they come as the public-relations battle over the use of animals in research, in Britain at least, seems to have been won. Violent direct action against scientists and laboratories has ended. Public sympathy and political support for the work have been secured. A key part of that victory was the ability to stress, time and time again, that the research was permitted only because it was always done under the strictest possible conditions. We cannot say that today. ■

The patent bargain

An open-source patent database highlights the need for more transparency worldwide.

It may take a patent lawyer to fully understand the scope of US patent number 7,777,022, but one thing is clear: at first glance, it certainly appears broad. The patent includes 4.2 million genetic sequences, some of which were identified computationally in a fishing trip for sequences that have applications in virology.

In June, the US Supreme Court determined that patents should no longer be granted for 'inventing' naturally occurring human genes, ending 30 years of the practice at the US Patent and Trademark Office. The decision will probably affect the growing genetic-diagnostics industry, and its influence will extend to patents on genes from other organisms. But it did not abolish all claims on DNA sequences — some

have estimated that the case will affect only about 8,000 of the at least 72,000 US patents that mention DNA sequences of one sort or another.

That leaves businesses with the unenviable task of sifting through the remainder to determine which, if any, will affect the commercialization of a given invention. Patent 7,777,022 highlights the growing difficulty in doing so: although it lists millions of sequences, it lays claim to only a few. A firehose of data and limited search tools make it impossible for all but highly trained patent specialists to make sense of the landscape around any technology. Highly trained patent experts do not come cheap: companies invest millions each year to keep track of the shifting intellectual-property landscape. Those that cannot afford the fee take the risk of being unable to patent their discoveries, or of being sued.

On 6 December, a study published in *Nature Biotechnology* took an important step towards rectifying that problem by revealing an open-source database that allows interested parties to map out the patent landscape around a technology without racking up exorbitant legal fees (O. A. Jefferson *et al.* *Nature Biotechnol.* **31**, 1086–1093; 2013).

The database, called the Lens (www.lens.org/lens), was created by Cambia, a non-profit organization in Canberra dedicated to facilitating innovation. It pulls together information from more than 90 patent jurisdictions worldwide. The Lens can be used to investigate patents of any ilk. But it has dedicated tools to analyse patents on DNA and protein sequences, and has plans to develop similar tools for other classes of patents, including those for circuits, software and chemicals.

The Lens is a bold effort to bring clarity and parity to the analysis of patents. It is also an innovation in need of support. Powered by eight busy software engineers, and funded by a patchwork of foundations and the Queensland University of Technology in Brisbane, Australia, it is tackling big-data problems that few have dared to take on. It will work best when it has cultivated a wiki-style following of users willing to take the time to annotate content, develop tools and share analyses.

Such a following can be hard to come by when academics and business leaders are already juggling busy schedules and scrambling for funding of their own. Cambia founder and chief executive Richard Jefferson is quick to acknowledge that some previous open-source efforts met with much enthusiasm but little participation from the academic community. It would be worth the effort for funders and institutions to find ways of incentivizing participation in an open-source patent effort.

Technology-transfer offices can help by logging the allocation of licences and changes in ownership in patent-assignment databases, where possible. A recent study led by Arti Rai, a specialist in intellectual-property law at Duke University in Durham, North Carolina, found that many universities fail to comply with basic requirements to acknowledge the contribution of federal funding to inventions in patent databases (A. Rai and B. Sampat *Nature Biotechnol.* **30**, 953–956; 2012).

“Many patent systems do not post their patents in a machine-readable format.”

Such information is important to track the history of the patent and the impact of federal research funding, as well as to allow the federal government to pursue its rights regarding such patents.

In the United States, the push to boost patent transparency has gained much-needed attention from on high. Revelations that some businesses, sometimes known as ‘patent trolls’, have been amassing large patent estates and using them to threaten other firms with litigation has caught the attention of the US Congress and the administration of President Barack Obama. Lawmakers are now considering legislation to rein in patent trolls, in part by creating reporting requirements that will help to clarify who owns a given patent — information that is currently hard to come by.

But the US patent system, troubled though it is, is not the only system that makes it difficult to track patents. In a survey published along with the Lens analysis, Cambia researchers noted that many patent systems do not routinely post their patents in a machine-readable format, making it difficult to search and analyse them. Where possible, it is time for such systems to address these flaws.

On the first day of many introductory patent-law classes, students are taught about the ‘patent bargain’. This is the foundation upon which the patent system is built: in exchange for protection for an invention, the inventor agrees to publicize their creation so that others may build upon it. The idea behind patenting was thus to put innovation into the public domain — yet the patent system has developed too many nooks and crannies in which information can be hidden away.

It is time to return to the bargain at the root of the patent system, and to use the computational and social-media tools at our disposal to publicize inventions, rather than obscure them. ■

Gender progress (?)

Despite some success, the proportions of women in Nature’s pages and as referees are still too low.

There are many obstacles to diversity in science. In any nation, there will be cultural and societal factors — often intersecting — that prevent the full research potential of one population group or another being fulfilled. One manifestation is discussed on page 211.

We at *Nature* have attempted to put our own house in order, and have produced just a scratch on the surface of one particular challenge — the low proportion of women contributing to our own content. That scratch is there thanks to actions taken since we focused on this issue in an Editorial a little over a year ago (see *Nature* **491**, 495; 2012).

So what have we achieved? In the visibility of women in our pages, progress has indeed been made.

In the News & Views section, the proportion of female authors has increased from 12% in 2011 to 19% in 2013.

The proportion of women appearing in profiles by our journalists has increased from 18% in 2011 to 40% in 2013. That does not include the four profiles in our ‘Women in science’ special issue early this year (see nature.com/women).

The number of articles by women in our World View section, which is driven by current topics, has remained low, now running at 12%. By contrast in 2013, 33% of Comment articles had at least one female author (27% of them had a woman as the first author). The combined total of World View and Comment articles with at least one woman author in 2013 is 26% — an improvement on 19% in 2011–12.

In our Editorial a year ago, we highlighted the need for a ‘gender loop’ — a conscious step in which an editor deliberately identifies several female candidates before selecting authors and profile subjects in our magazine sections, and referees for our research papers.

In this last category, the result has been disappointing — the number of women referees has remained all too low. From 14% in 2011, the proportion of women fell to 12% in 2012 and then rose to 13% in 2013. Taking into account uncertainties resulting from ambiguities in some names, these numbers are essentially on a plateau.

Efforts have been made by research editors, when visiting labs and meetings, and when surveying the literature, to increase the number of women invited to act as reviewers. Women already make up only a small proportion of the potential referees, owing to the demographics of the research community. And our efforts have made us all the more aware that a higher proportion of women than men decline our invitations to referee. We have not investigated this with a survey, but informal comments indicate that women tend to be that much more busy.

The lesson in this tale is: we must try harder. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhuncq



Preserve the European Research Council's legacy

A change in European Commission management styles under Horizon 2020 must not end the funding agency's focus on excellence, says Helga Nowotny.

Influential German sociologist Max Weber said that the 'routinization of charisma' is normal in organizations. It is fashionable these days to compare science with business, and Steve Jobs and Apple come to mind as examples: the company's image became based around the personal characteristics and accomplishments of its innovative, edgy and revolutionary leader, and so survived him.

I will soon step down as president of the European Research Council (ERC), and my successor will be announced next week. But when it comes to the succession plan of the ERC, the charisma comes not from a single person, but from the pioneering scientists who acted as the driving force to reshape the entire research landscape of Europe.

There is turnover, too, in the Scientific Council that steers the ERC strategy. By the end of next year, its last two founding members will have left. Although this is part of the normal renewal process, it must not lead to a loss of the ERC's institutional memory of its crucial founding years, including how it overcame many obstacles.

I am concerned for the future of the ERC — but first, some history. Every successful organization begins with a courageous decision, according to management consultant Peter Drucker. That applies to the ERC. Seven years ago, the European Union (EU) made a courageous decision to establish an agency to fund frontier research on the basis of a sole criterion: scientific excellence. It received a 'big bang' budget of €7.5 billion (US\$10.3 billion). Most unusually, setting the scientific strategy was entrusted to the autonomous Scientific Council of 22 members, whereas the implementation would be in the hands of a 'dedicated implementation structure', which is now the ERC Executive Agency.

The ERC has been a great success. Recognition of the fairness, transparency and credibility of its evaluation procedure has transformed it into the gold standard for Europe. This high reputation has unleashed unprecedented competition among host institutions for the prestige of ERC grants. This has probably contributed more than anything to energizing the European Research Area and increasing its attractiveness to researchers from outside Europe. Most importantly, it has boosted the careers of the younger researchers who now constitute almost two-thirds of the more than 4,000 grant recipients whom the ERC has funded so far.

As these numbers grow, so does the number of people involved in ERC evaluations, from 900 panel members in 2009–10 to almost 2,000 in 2012–13. But the scientific community should not think that ERC activities have become a safe, predictable routine.

Here is my concern. Neither the ERC nor the standards it has set for scientific excellence

should be taken for granted. At stake is the continuing sense that the ERC belongs to the scientific community, earned through scientists' continued commitment and dedication as panel members, remote reviewers, constructive critics and mentors. The founding generation's legacy is an evaluation procedure of the highest quality, and it must be maintained and safeguarded at all costs.

Under Horizon 2020, the next round of EU funding for research and innovation, the ERC will have a significantly increased budget of more than €13 billion. This is an explicit recognition by the European Commission of the pioneering part that the ERC is expected to play in the next seven years, injecting scientific excellence into other parts of Horizon 2020 and demonstrating the impact of frontier research.

This is good news. But the consequences of some of the changes under Horizon 2020 are less clear. The ERC's dual governance structure is both its unique strength and its greatest vulnerability. The strength is the power of its autonomous Scientific Council. The vulnerability is that implementation of the council's decisions by the ERC Executive Agency is subject to general EU rules and regulations, not all of them fit for the ERC.

After overcoming many initial difficulties, the Scientific Council has forged a trusted relationship with its Executive Agency and a highly professional staff. Under the new management modes of Horizon 2020, there will be a massive outsourcing of activities to executive agencies; the ERC Executive Agency will become one of many. I can see the pressure mounting for

administrators to comply with a 'one size fits all' model based on the common denominator of executive agencies and their underlying logic of greater efficiency through streamlining. New key performance indicators, such as impact measurements, are being set up, and some may be at odds with the ERC's mission of funding only excellence.

The European Commission will continue to oversee and control the executive agencies. The special nature of the ERC is enshrined in EU legislation. The Horizon 2020 big machinery will soon start working, but it must not be allowed to overrun the unique specificity of the ERC by subjecting it to rules that do not fit its objectives. The commission must guarantee that the ERC is not put into an administrative straitjacket.

The ERC sees itself as a learning, and hence experimental, institution. It has proven that a courageous decision such as that by those who established it can and does pay off. Continuous vigilance, monitoring and adjustments will be necessary to safeguard what has been accomplished. ■

Helga Nowotny is president of the European Research Council. Her term ends on 31 December.
e-mail: helga.nowotny@wwtf.at

**THE COMMISSION MUST
GUARANTEE
THAT THE ERC IS NOT
PUT INTO AN
ADMINISTRATIVE
STRAITJACKET.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/zywhwj

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

AGROECOLOGY

Bees are better for strawberries

Bee pollination doesn't just boost yields of fruit crops — it also improves fruit quality, at least in strawberries.

Strawberry plants can self-pollinate, or be pollinated by wind or bees. Björn Klatt of the University of Göttingen in Germany and his colleagues grew the fruit using a permeable plastic to eliminate wind and bee pollination, or fine-mesh bags to exclude just bees. They found that berries grown uncovered but pollinated by bees were heavier, redder, and had a longer shelf life and a more desirable sugar-to-acid ratio than wind- or self-pollinated berries.

The researchers say that bee pollination boosts hormonal growth regulators that improve the quality of the berries.

Proc. R. Soc. B <http://doi.org/qcg> (2013)

CLIMATE CHANGE

Melting ice spurs wild weather

Recent weather extremes during summer in the Northern Hemisphere, such as the 2012 drought in the United States (**pictured**), seem to be linked to loss of Arctic sea ice and reductions in snow cover.

Qihong Tang of the Institute of Geographic



Sciences and Natural Resources Research in Beijing and his colleagues compared meteorological records with satellite observations of snow and ice. They found that Arctic sea-ice retreat and, to a lesser degree, decreased snow cover alter large-scale atmospheric circulation patterns — for example, by shifting the jet stream northwards. These changes typically cause heatwaves and other extreme weather events at mid-latitudes.

The results could help to improve seasonal and longer-term climate forecasts, the

team says.

Nature Clim. Change <http://doi.org/qds> (2013)

ASTROPHYSICS

Trio of distant quasars found

Astronomers have discovered three distant quasars that will allow them to probe the conditions of the early Universe.

A team led by Bram Venemans at the Max Planck Institute for Astronomy in Heidelberg, Germany,

or turned away from the animals.

Baboons made more begging hand gestures when experimenters had their eyes open. When the experimenters' eyes were closed or their backs were turned, the monkeys tended to make noisy banging actions instead. Attention-getting gestures in monkeys may have been a starting point for the evolution of intentional vocal communication in great apes and humans, the researchers suggest.

Anim. Behav. <http://doi.org/qcb> (2013)



ANIMAL BEHAVIOUR

Baboons know when to be noisy

Baboons can tell whether they need to make a ruckus to get attention or if they already have a captive audience.

To see whether the animals adjust their mode of communication according to the level of an observer's attention, Marie Bourjade at Aix-Marseille University, France, and her colleagues tempted 16 captive olive baboons (*Papio anubis*; **pictured**) with a tasty fruit treat. The experimenters stood facing the animals with their eyes either open or closed,

discovered the active galactic nuclei using the European Southern Observatory's Visible and Infrared Survey Telescope for Astronomy Kilo-degree Infrared Galaxy (VIKING) survey. The findings bring the number of known quasars that are beyond the detection limit of visible-light cameras to four.

The authors say that studying the quasars, which are thought to be powered by supermassive black holes more than 1 billion times the mass of the Sun, will shed light on how massive galaxies and black holes formed

ANUP SHAH/CORBIS

VICTOR J. BLUE/BLOOMBERG VIA GETTY

less than 900 million years after the Big Bang.

Astrophys. J. 779, 24 (2013)

BIOTECHNOLOGY

CRISPR corrects genetic disease

A system for editing genes has now been used to repair disease-related mutations in mice and human stem cells, highlighting the technology's therapeutic potential.

The recently developed CRISPR system uses an RNA strand matching a target gene to guide a bacterial enzyme, Cas9, to excise the gene. Jinsong Li and his colleagues at the Shanghai Institutes for Biological Sciences in China used the technique in fertilized mouse eggs to correct the mutated *Crygc* gene, which causes cataracts. The mice grew into healthy adults that bore normal offspring.

In a separate paper, Hans Clevers at the Hubrecht Institute in Utrecht, the Netherlands, and his team repaired the cystic fibrosis gene, *CFTR*, in cultured intestinal stem cells obtained from patients with the disease. *Cell Stem Cell* 13, 653–658; 659–662 (2013)

METEOROLOGY

Satellite improves storm forecasts

Data from a US Earth-observing satellite could help improve the accuracy of predictions of hurricane track and strength.

When generating hurricane forecasts, the US National Weather Service does not use real-time information from weather satellites. But Xiaolei Zou at Florida State University in Tallahassee and her colleagues looked at the effect of including data from the

Suomi NPP satellite, launched in 2011, on hurricane forecasts. The satellite's microwave instrument measures air temperature and humidity.

Incorporating Suomi data into the government's hurricane model for four 2012 storms, including Sandy (pictured), made for more accurate forecasts of track and intensity. The work suggests a way to improve the notoriously difficult predictions of storm strength. *J. Geophys. Res. Atm.* 118, 11558–11576 (2013)

NEUROSCIENCE

Primate brain makes oestrogen

Ovaries are the main producers of oestrogen but when they give out, the brain in monkeys can take over in making the hormone.

Ei Terasawa and her colleagues at the University of Wisconsin in Madison removed the ovaries of female rhesus monkeys (*Macaca mulatta*) and briefly stimulated the animals' brains with oestrogen or an electrical current. The brain's hypothalamus region began producing its own oestrogen, along with other hormones involved in reproduction.

If human brains have the same capability, targeting this oestrogen source could be a way to treat diseases such as depression that may be linked to oestrogen imbalances, the authors say.

J. Neurosci. 33, 19051–19059 (2013)

ECOLOGY

Why rabies hangs on after bat culls

Culling vampire bats in South America to combat the spread of rabies could be having the opposite effect, according to a modelling study.

Vampire bats (*Desmodus rotundus*) transmit the rabies virus by biting humans and livestock.

COMMUNITY CHOICE

The most viewed papers in science

ENVIRONMENTAL SCIENCE

Gas production contaminates water

HIGHLY READ
on pubs.acs.org
in November

Waste water from oil and gas production is contaminating surface waters and sediments in western Pennsylvania — even after it has been treated.

In the United States, such waste is sometimes sent to treatment facilities and then discharged into streams and rivers. Nathaniel Warner, Avner Vengosh and their colleagues at Duke University in Durham, North Carolina, analysed the effluent in 2010–12 at a facility that treated water from gas wells in the Marcellus Shale. They looked at samples taken upstream and downstream of the facility, and found increased contaminants such as chloride and bromide downstream.

Although treatment lowered the levels of barium and radium in the waste water, radioactivity levels from radium in river sediments near the facility were 200 times greater than background levels, and were above regulatory limits.

The authors say that the elevated radium levels suggest a risk of radioactivity accumulating in wastewater disposal areas, and that better treatment technologies are needed to reduce contamination.

Environ. Sci. Technol. 47, 11849–11857 (2013)

Julie Blackwood at Williams College in Williamstown, Massachusetts, and her colleagues analysed field data on individual bats collected from 17 colonies in Peru between 2007 and 2010. They found that the rabies virus probably does not persist in a single bat colony, but instead is spread between colonies by wandering bats. Most infections do not kill bats but confer temporary immunity, which also maintains the levels of virus.

Bat culls, the researchers note, may increase movement of bats and thus promote the spread of the disease.

Proc. Natl Acad. Sci. USA
<http://doi.org/qb9> (2013)

stomachs have been found in ancient oceans and lakes.

David Hone, at Queen Mary University in London, and Donald Henderson at the Royal Tyrrell Museum of Palaeontology in Drumheller, Alberta, tested the floating-pterosaur hypothesis by creating computer models of what four pterosaurs might have looked like afloat. They took account of the reptiles' bone density, skeleton shape and other factors. The models suggest that, under most conditions, pterosaurs would have oriented their heads horizontally to the water, leaving the bottom one-quarter to one-third of their heads submerged.

Pterosaurs would have risked drowning if they spent too much time on the water, the researchers say.

Palaeogeogr. Palaeoclimatol. Palaeoecol. <http://doi.org/qcf> (2013)

PALAEOLOGY

Ancient reptiles stuck to the air

Flying reptiles that went extinct around 65 million years ago probably did not spend much time on water — even though fossils of the creatures with fish in their

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

Dementia cases

The number of people living with dementia around the world is now estimated to be 44 million, up 22% from three years ago, according to a report released on 5 December by Alzheimer's Disease International (ADI), a worldwide federation of Alzheimer's disease associations. As people live longer, cases of dementia will triple to 136 million by 2050, the ADI says. See go.nature.com/ikdput for more.

Palm-oil pledge

Singapore-based agribusiness giant Wilmar International announced on 11 December that it would no longer operate or purchase from palm plantations made on recently deforested land. Palm oil is used in many food products and to produce biofuels, and the expansion of palm plantations has driven massive deforestation in places such as Indonesia and Malaysia. Wilmar is the world's largest palm-oil supplier.

UK spending

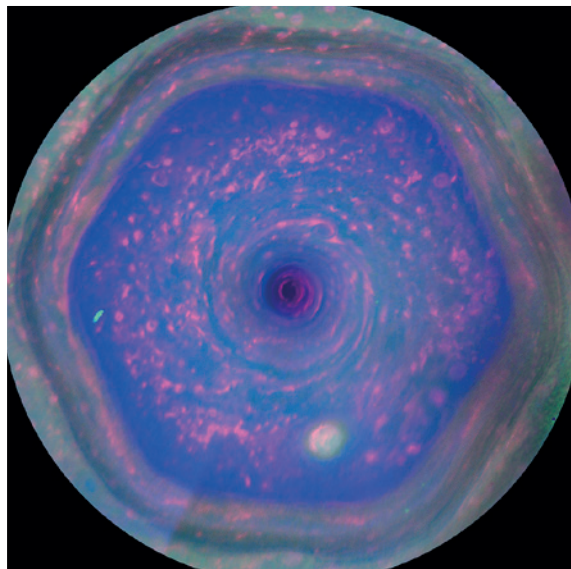
The UK government will spend £270 million (US\$441 million) over the next five years on a network of centres for quantum technology, Chancellor of the

NUMBER CRUNCH

-94 °C

The coldest temperature ever recorded on Earth. NASA satellites pinpointed the low in Antarctica, where the previous record of -89.2 °C was measured in 1983.

Source: NSIDC



Saturn's six-sided storm

NASA last week released high-resolution videos of a hexagonal jet-stream pattern, about 30,000 kilometres across, around the north pole of the gas giant Saturn. The images, snapped by the Cassini spacecraft, show in detail the stream of 320-kilometre-per-hour winds that was first spotted by the Voyager probe in the early 1980s.

Exchequer George Osborne announced last week. In an annual budget statement, he added that the country would remove caps on student numbers at universities (allowing more money to flow to science courses), and spend £375 million over five years in building research capacity in collaboration with 'emerging powers' such as China. See go.nature.com/wbxama for more.

Animal welfare

The treatment of laboratory animals at one of the United Kingdom's most prestigious universities has been severely criticized by an independent review, set up in the wake of an undercover investigation by an animal-rights group. Imperial College London's

animal-research facilities are under-resourced and operating without adequate systems for supervision and ethical review, found the report, which was published on 10 December. See page 187 and go.nature.com/4dxlf4 for more.

Planetary hole

Researchers at NASA's planetary science division are grumbling that a reshuffle could see them miss out on grants next year. Under a newly formed programme, announced on 3 December, research funding has been sorted into five large themes — but the biggest area, Solar System workings, will not be calling for funding proposals until 2015. The gap is worrying some scientists who are used to living from grant

to grant — and are already concerned by shrinking budgets. See go.nature.com/ntui1n for more.

Global Fund cash

The Global Fund to Fight AIDS, Tuberculosis and Malaria secured US\$12 billion in firm funding pledges for the next three years at a meeting with donors in Washington DC last week. The amount was more than the \$9.2 billion in firm pledges that the fund, based in Geneva, had obtained in 2010 for the period 2011–13. But it fell short of the \$15 billion the fund sought.

RESEARCH

HIV hope dashed

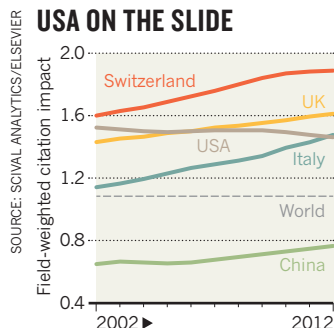
HIV has reappeared in two people who had seemed to be free of the virus after receiving bone-marrow transplants. The two men were treated in 2008 and 2010, respectively, and in July doctors said that the pair had remained free of HIV for eight months after they stopped taking antiretroviral medication. But the virus has returned, researchers said last week; a disappointing but not surprising result given that the transplanted cells were not resistant to HIV (unlike in the case of the successfully cured 'Berlin patient', Timothy Ray Brown). See go.nature.com/vy2eas for more.

US quality ranking

The United States is sliding down the ranks of research quality as measured by the relative citation impact of its papers, according to a study commissioned by the UK government (see **graph**). It was overtaken in the rankings (normalized for field) by the United Kingdom in 2006 and by Italy in 2012, say figures from SciVal Analytics, the analysis arm of publishing firm Elsevier. But the United

NASA/JPL-CALTECH/SSI/HAMPTON UNIV.

USA ON THE SLIDE



States remains well ahead in terms of its share of the world's top 1% cited articles.

Bacterial genomes

DNA sequencing company Pacific Biosciences has announced an effort to fully sequence the genomes of 3,000 bacterial strains over three years. The company, based in Menlo Park, California, said on 3 December that it would collaborate with the Wellcome Trust Sanger Institute, near Cambridge, UK, and Public Health England, whose National Collection of Type Cultures will provide the bacterial strains. The 93-year-old collection contains more than 5,000 bacterial strains, including many human and animal pathogens.

PEOPLE

Polar-bear payout

A federal scientist has won a \$100,000 payout from the US Department of the

Interior, which had accused him of publishing false data in a paper reporting that four polar bears had drowned in the Arctic (C. Monnett and J. S. Gleason *Polar Biol.* **29**, 681–687). Charles Monnett was cleared of any misconduct last year after a government investigation (see go.nature.com/dfio5o), and he subsequently filed a whistleblower complaint. In the settlement, announced on 4 December, Monnett also agreed to retire from his post at the Bureau of Ocean Energy Management in Washington DC.

Data chief

The US National Institutes of Health (NIH) is getting into big data in a big way. On 9 December, the NIH named pharmacologist Philip Bourne of the University of California, San Diego, as its first permanent associate director for data science. Bourne's office will handle NIH-wide initiatives involving big data from fields such as genomics and electronic medical records. He will start at the NIH in 2014.

Criminal suspect

An ongoing criminal investigation into the state-funded Cancer Prevention and Research Institute of Texas (CPRIT) has nabbed a suspect. On 3 December, a Texas attorney general

filed an indictment against Jerald Cobbs, former chief commercialization officer for CPRIT in Austin, for allegedly awarding an US\$11-million grant to the Dallas-based drug company Pelotron Therapeutics without approval from the proper review committee. For more on concerns at CPRIT, see *Nature* **486**, 169–171 (2012).

BUSINESS

Start-up stimulus

A US\$100-million initiative to support life-sciences start-ups in New York City was launched on 4 December by the New York City Economic Development Corporation. The non-profit organization will provide \$10 million, with \$50 million coming from venture capitalists. Three firms will offer the remaining \$40 million: Eli Lilly in Indianapolis, Indiana; Celgene in Summit, New Jersey; and GE Ventures in Fairfield, Connecticut.

Gene testing

Personal-genomics company 23andMe will no longer provide health-related information to new customers, the firm said on 5 December. The move came in response to a 22 November warning letter from the US Food and Drug Administration (FDA) telling the company, which is based

COMING UP

14 DECEMBER

China's Chang'e-3 spacecraft attempts to land on the Moon.

in Mountain View, California, to stop marketing its genetic-testing service as a medical device without approval (see *Nature* <http://doi.org/qdm>; 2013). A class-action lawsuit, filed against the firm on 27 November in the Southern District of California, accuses it of false advertising for selling its saliva-testing kit without the FDA's blessing.

Clinical-trial data

Drug firm Pfizer has joined companies that are opening up access to their clinical-trial information (see *Nature* **495**, 419–420; 2013). The company, which is headquartered in New York, said that it would consider requests for anonymized patient-level data from trials completed more than two years ago if the trial products were approved or discontinued. Summaries of the 'clinical study reports' (CSRs) for approved products will also be published online. The AllTrials group, which campaigns for more access to clinical-trial data, said that the statement contained loopholes — such as excluding 'off-label' studies — and that full CSRs should be available.

Pricey pill

On 6 December, the US Food and Drug Administration approved sofosbuvir, a pill for hepatitis C made by Gilead Sciences in Foster City, California. Unlike older, injected therapies, sofosbuvir can be administered without being accompanied by interferon, which has toxic side effects. The course of treatment will cost \$84,000 — \$1,000 per pill.

► **NATURE.COM**

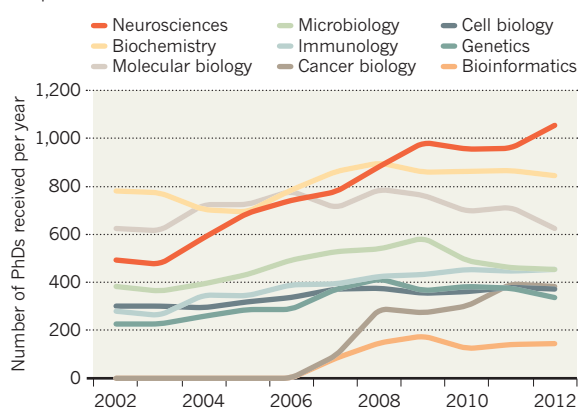
For daily news updates see:
www.nature.com/news

TREND WATCH

The latest release of the US National Science Foundation's survey of earned doctorates in the United States reveals the remarkable rise of PhDs in neuroscience over the past decade compared with other biological disciplines. Doctorates in that field have more than doubled, whereas the total number of doctorates has increased by only 27%. Other fast-growing disciplines include bioengineering (382% rise) and computer sciences (228%).

MORE BRAINS STUDYING BRAINS

The number of US PhDs awarded in neuroscience has doubled over the past decade.



NEWS IN FOCUS

GEOLOGY Mapping project opens way to sinkhole predictions **p.196**

AGRICULTURE Fungi offer fresh take on additives to improve crops **p.199**

PHYSICS Big Bang afterglow may have offered right conditions for life **p.201**



MEDICINE Radical recipes to beat radioisotope shortage **p.202**

SANJIT DAS/BLOOMBERG/GETTY



Banana plantations are at risk from a disease known as *Fusarium* wilt.

AGRICULTURE

Fungus threatens top banana

Fears rise for Latin American industry as devastating disease hits leading variety in Africa and Middle East.

BY DECLAN BUTLER

A variant of a fungus that rots and kills the main variety of export banana has been found in plantations in Mozambique and Jordan, raising fears that it could spread to major producers and decimate supplies. The pathogen, which was until now limited to parts of Asia and a region of Australia, has a particularly devastating effect on the popular Cavendish cultivar, which accounts for almost all of the multibillion-dollar banana

export trade. Expansion of the disease worldwide could be disastrous, say researchers.

The disease is caused by strains of a soil fungus called *Fusarium oxysporum* f. sp. *cubense* (Foc). A strain of Foc previously wiped out the Gros Michel cultivar, which was the main exported banana variety from the nineteenth century until the 1950s. In response, the industry replaced Gros Michel plants with the Cavendish variety, which is resistant to that Foc strain. But Cavendish is susceptible to the new Foc Tropical Race 4 (Foc-TR4) strain, and

could meet the same fate as Gros Michel if the fungus reaches Latin America, the world's leading banana exporter, says Rony Swennen of the Catholic University of Leuven in Belgium, and a banana breeder at the International Institute of Tropical Agriculture in Dar es Salaam. "It's a gigantic problem," he adds. Although Foc strains spread slowly, they are almost impossible to eliminate from soil.

Foc-TR4 was first detected in Asia in the 1990s, and is now found in Taiwan, Indonesia, Malaysia, the Philippines, China and northern Australia (see 'Fruit threat'). The outbreak in Jordan, reported on 29 October (F. A. Garcia *et al. Plant Dis.* <http://doi.org/qd3>; 2013), was the first to be described outside those nations. The Mozambique outbreak was reported last month.

Nobody is sure how the fungus arrived in Jordan or Mozambique. Migrant workers from Asia might inadvertently have brought contaminated soil with them. Another possibility is the import of infected rhizomes — the stems from which banana plants propagate. But much of the Cavendish industry now uses tissue culture, which produces pathogen-free plantlets.

To slow the spread, good farm hygiene, and prompt quarantine and destruction of infected plants are crucial. Altus Viljoen, a researcher at Stellenbosch University in South Africa, was called in to identify the cause of the Mozambique outbreak, and says that authorities were quick to take action. He estimates that the disease has been present for two to three years.

Gert Kema, a *Fusarium* researcher at Wageningen University and Research Centre in the Netherlands and co-author of the Jordan report, believes that further spread is inevitable. "I'm incredibly concerned," he says. "I will not be surprised if it pops up in Latin America in the near future." That region, along with the Caribbean, accounts for more than 80% of banana exports. If Foc-TR4 takes root there, it could lead to the slow demise of industrial farming of the Cavendish variety.

Smaller farms in Asia are already trying to mitigate losses. Tissue culture of Cavendish plants has generated variants with random mutations that confer partial resistance to Foc-TR4. Planting of these variants, in combination with measures such as crop rotation, has allowed the cultivation of bananas on contaminated land. But production losses and higher costs make affected plantations less economically viable.

Progress in creating bananas fully resistant ▶

► to Foc-TR4, either by classical breeding or genetic engineering, has so far been limited. The wild Asian banana *Musa acuminata malaccensis* — the genome of which was published last year (A. D'Hont *Nature* **488**, 213–217; 2012) — seems to be resistant, and researchers are experimenting with putting its resistance genes into the Cavendish. The resulting transgenic specimens have been in field trials for 18 months on contaminated ground in Australia, and are looking “very promising”, says James Dale, director of the Centre for Tropical Crops and Biocommodities at Queensland University of Technology in Brisbane, Australia. But he cautions that the full results are not yet in.

For those who buy their bananas in supermarkets, the Cavendish may well be the only variety they know. But exports of the cultivar account for only about 13% of the 150 million or so tonnes of bananas and cooking bananas (plantains) produced annually. Industrial farms growing a single Cavendish cultivar are at a high risk of Foc-TR4 infestation, but the fungus poses less of a threat to the bulk of the bananas that provide a staple for some 400 million people worldwide.

Most of the bananas important for the food supply are grown by smallholder farmers in low-income countries and consumed locally. Hundreds of cultivars are farmed, and this biodiversity is an important rampart against disease. Researchers do not yet have a full picture of the susceptibility of these varieties, but many cultivars are likely to be resistant to Foc-TR4 because they are biologically different to the Cavendish.

For his part, Dale is trying to engineer Gros Michel bananas for resistance to the original Foc strain. The Cavendish is bland by comparison and it bruises more easily. Dale would like to see Gros Michel on supermarket shelves again. “It’s such a superior banana to Cavendish. To bring it back would be wonderful.” ■ [SEE CORRESPONDENCE P.218](#)



Grounds for concern: a sinkhole in Florida devoured a pool, a boat and two houses on 14 November.

GEOLOGY

Florida forecasts sinkhole burden

Predictive model will map areas vulnerable to collapse.

BY ALEXANDRA WITZE

A 21-metre-wide hole in the ground opened up behind two homes last month in Dunedin, Florida, swallowing a swimming pool, a boat and eventually both houses. Days later, geologist Clint Kromhout found himself on the phone with a woman who lived nearby. She was upset, and ready to sell her house and leave the state.

Kromhout, who works with the Florida Geological Survey in Tallahassee, gave her his usual speech: anyone here should get used to sinkholes. In the easily erodible karst terrain of Florida, they strike regularly and seemingly at random. In August, sinkholes collapsed a three-storey holiday villa near Disney World. In February, a sinkhole in Seffner entombed a man while he slept. “If you live in Florida, a bunch of things are given: sunshine, beaches, hurricanes — and sinkholes,” Kromhout says.

But this autumn, he and his colleague Alan Baker began to develop a weapon in the battle against their state’s geology: a predictive map. By 2016, they expect to have a state-wide map that is colour-coded by category of sinkhole vulnerability at a scale of around 1 kilometre. “We want to produce a map that is scientifically defensible and not just based on some expert’s opinion,” says Kromhout.

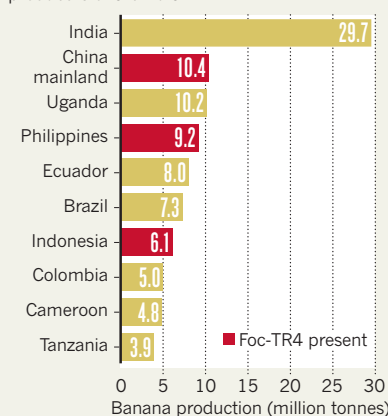
The result will feed into the multimillion-dollar battle between homeowners and insurance companies in Florida, where coverage for ‘catastrophic ground cover collapse’ is required but damage claims are often denied. It could also set a new standard for sinkhole-vulnerability studies in other states and countries, which typically compile maps of existing sinkholes but do not have predictive maps showing where new ones might form. Carbonate rocks such as limestone, which are prone to being chewed up into

LUKE JOHNSON/THE TAMPA TRIBUNE/AP

SOURCE: FAO

FRUIT THREAT

A fungus strain that kills banana plants has been detected in three of the world’s top producers of the fruit.



Swiss-cheese-like karst terrain, cover 13% of the planet's land (see 'A sinking feeling'), and the problems that they create plague not only Florida but also much of the United States, Europe, China and beyond.

How sinkholes form is well known. In one common scenario, naturally acidic rainwater percolates through carbonate rock and dissolves it to form caves and other underground openings. Overlying sediments then collapse into the void, forming a sinkhole.

Florida is particularly vulnerable because its layers of limestone are at most only a few tens of millions of years old; they have not had time to become compressed and more resistant to dissolution and erosion. In places, interleaved layers of sand from ancient beaches enhance the erosion of limestone. Layers of clay near the surface are strong enough to hold sediments together even as voids form beneath them, but when the clays do give way, the chasms can be immense. Drought, hurricanes and the pumping of groundwater cause fluctuations in the water table — a further factor in the state's propensity for sinkholes. "There are different styles and flavours of karst, but Florida is an animal all its own," says Daniel Doctor, a karst expert at the US Geological Survey in Reston, Virginia.

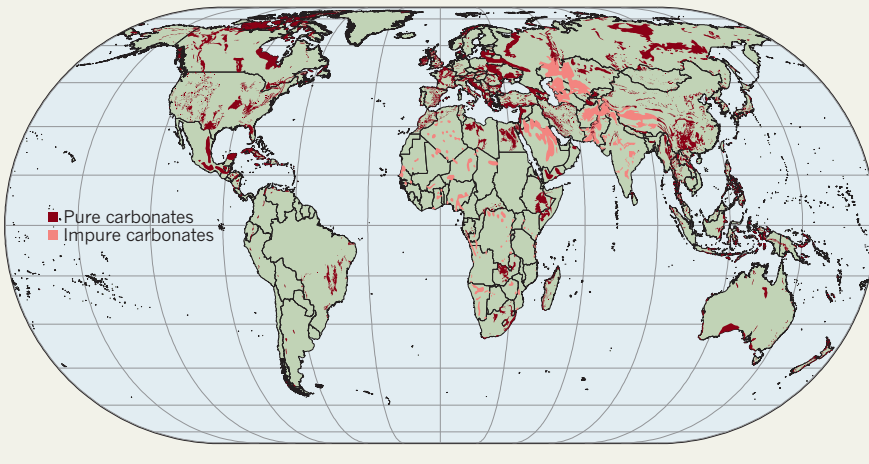
In June 2012, the issue came to the fore after Tropical Storm Debby swept across the state, inundating the ground after months of drought and causing hundreds of sinkholes to appear. That's when the state division of emergency management contacted the Florida Geological Survey to see what it could do to prepare for future outbreaks.

The state survey has mapped sinkholes for more than a century, but the records are patchy and biased towards urban and suburban areas, where sinkholes are more likely to cause property damage when they occur. The survey keeps a list of almost 3,500 reported depressions, but some of those are unlikely to be 'proper' sinkholes, says Jonathan Arthur, director of the state survey. Many could be other types of collapses caused by buried wood piles or old septic tanks that caved in.

The sinkhole maps are also, by definition, retrospective. But, funded by almost US\$1.1 million from the state and the Federal Emergency Management Agency, Kromhout and Baker are using a technique that can

A SINKING FEELING

Carbonate rocks cover 13% of the world's land surface and are susceptible to erosion by running water. The resulting karst terrain tends to have high numbers of caves and sinkholes.



predict future occurrences. Called 'weights of evidence', the three-decade-old method has its roots in the diagnosis of medical illnesses, and it has also been used in environmental settings such as mineral exploration. First, researchers collect information that is already held in various databases — factors such as the depth of limestone and the water table, and the presence of clays and streams. A computer model integrates the information and spits out a prediction of where sinkholes are most likely to form. Then the sites of known sinkholes are used to test the model's predictions.

"There are different styles and flavours of karst, but Florida is an animal all its own."

At 1-kilometre resolution, the map will still not be able to tell a property owner if a hole is likely to appear below his or her shed. But it should prove helpful to emergency managers who are seeking clues as to where the next sinkhole swarm might form. Insurance companies are also likely to combine the map with their own studies when setting insurance premiums in different regions, Arthur says.

Officials at all levels of government are looking to incorporate more solid scientific information into sinkhole policies, says Robert Denton, a senior geologist with

GeoConcepts Engineering in Ashburn, Virginia. For instance, Virginia is looking to adopt standardized guidelines that any expert hired to assess a possible sinkhole would have to follow. And next year, the US Geological Survey is expected to release the first digitized version of its National Karst Map — a step that could push other states to develop predictive maps.

Across the globe, other governments have embarked on related steps. The Vaud canton in western Switzerland has been mapping sinkholes within its borders, in part to improve insurance requirements on public buildings in case of sinkhole damage. In Italy, scientists at the Institute of Research for Hydrogeological Protection (IRPI) in Bari have been compiling a database of sinkhole distribution.

The Italian catalogue, which currently includes 900 sinkholes, is the first there to incorporate time of occurrence as well as location, says Mario Parise of the IRPI. That's crucial for planning, he says, because local officials want to know not only where a sinkhole might form but also when.

For the Florida team, plenty of work remains. This week, the researchers plan to explore state parks, checking to see whether circular features shown on maps really are sinkholes. From there, it's a matter of hard work, then waiting to see whether they can predict the next hole in the ground. ■


**MORE
ONLINE**

TOP STORY



Thin clay fault caused by earthquake and tsunami.
go.nature.com/t2zkdY

MORE NEWS

- Dyslexia linked to miscommunication between brain centres go.nature.com/mpba3p
- Large-scale solar convection cells confirmed go.nature.com/jecc7m
- Mortality and fertility mapped across phyla go.nature.com/hz7oy7

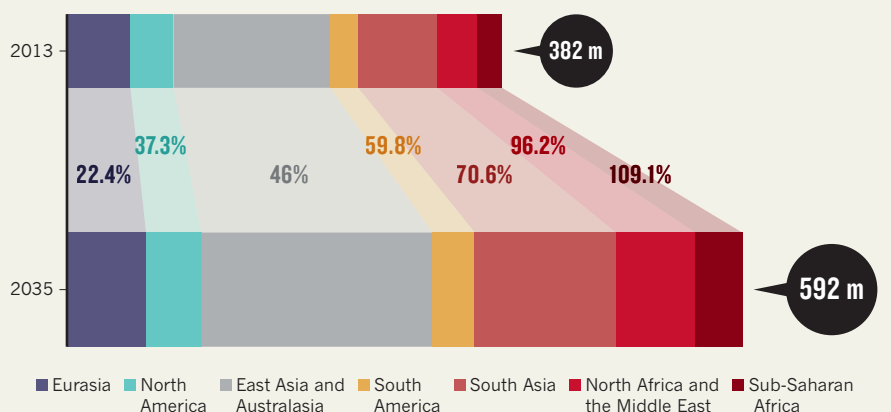
NATURE PODCAST



Lifespans across the tree of life; lizards breathe like birds; and the medical-isotopes crisis nature.com/nature/podcast

A GLOBAL EPIDEMIC

By 2035, 592 million people will have diabetes, with the largest increases (%) occurring in the developing world.



MEDICINES

Diabetes drugs ride a bumpy road

Safety worries hamper emerging therapies.

BY HEIDI LEDFORD

Biochemist Matthias Hediger immediately saw the possibilities when his team discovered the protein SGLT2, which is used by the kidneys to soak up sugar from urine. Blocking the protein, he reasoned, would boost the loss of sugar from the body — a boon for diabetics with skyrocketing blood glucose levels. Furthermore, the mechanism would be independent of insulin, the hormone that regulates the absorption of sugar from the blood and the target for most diabetes drugs. “It was obvious it was important,” says Hediger, now at the University of Bern.

Pharmaceutical companies have shared Hediger’s enthusiasm in the 20 years since that discovery: a dozen are working to develop SGLT2 inhibitors. The drugs are expected to earn US\$6 billion of the \$47-billion global market projected for type 2 diabetes medications by 2022, according to Decision Resources, a market-research company based in Burlington, Massachusetts.

But the drugs must first get past a regulatory agency that is on high alert. The list of safety risks for diabetes drugs keeps growing, following scares over liver and cardiovascular side effects of some drugs. Several worries have enveloped SGLT2 inhibitors, even a promising drug called dapagliflozin, which faces a hearing before advisers at the US Food and Drug Administration (FDA) on 12 December.

“Safety issues have been a major cloud in

diabetes drug development recently,” says Andrew Krentz, an endocrinologist at the Profil Institute for Clinical Research in Chula Vista, California, which specializes in diabetes and obesity. “The scrutiny has been particularly intense.”

Worldwide, 382 million people have diabetes, a number that by 2035 is expected to climb to 592 million — almost 10% of the world’s present population, according to the Brussels-based International Diabetes Federation (see ‘A global epidemic’). That growth is driving the pharmaceutical frenzy, but it also amplifies safety concerns: with such a large patient base, even rare side effects can affect a significant number of people, especially given that drugs for this chronic disease must be taken for years. Also, patients with type 2 diabetes tend to be obese and are prone to a host of other conditions, including heart disease and some cancers.

Since 2008, the FDA has required drug developers to conduct clinical trials to test the heart risks of new diabetes medications. The decision came in the wake of a scare over Avandia (rosiglitazone), a popular diabetes drug that boosts sensitivity to insulin and that became linked to heart failure and stroke eight years after it was approved for the market. Last month, the FDA exonerated Avandia, which is made by London-based GlaxoSmithKline, and

“The scrutiny has been particularly intense.”

rolled back restrictions on the drug’s use — but the experience, along with liver damage associated with a related medication, has left drug developers on edge.

Cardiovascular trials are expensive because they must enrol many patients and follow them for years to accrue enough heart attacks and strokes for statistical significance. Critics of the requirement fretted that it would stifle diabetes drug development. And that fear was bolstered in November when New York drugmaker Bristol-Myers Squibb announced that it was cutting its diabetes research efforts. But Sanjay Kaul, a cardiologist at Cedars-Sinai Medical Center in Los Angeles, California, says that since the 2008 FDA rule, companies have set up about 15 cardiovascular trials of diabetes drugs, involving some 115,000 patients. “To me, that is not a reflection of innovation being stifled,” he says.

The SGLT2 inhibitors have raised different concerns. In 2011, FDA advisers expressed worry over dapagliflozin, then leading the pack of SGLT2 inhibitors en route to market, after clinical trial data showed that patients taking the drug were more likely to develop breast or bladder cancer. The numbers were too small to reach statistical significance, but they were enough to convince the FDA to reject the drug in January 2012.

In July, the drug’s makers, Bristol-Myers Squibb and London-based AstraZeneca, submitted a new application with updated clinical-trial data, which will be discussed at this week’s FDA meeting. Although the nature of the data is not known, advocates point to several hopeful developments: in March, the FDA approved the first SGLT2 inhibitor, canagliflozin, and in April 2012, the European Medicines Agency approved dapagliflozin, despite the cancer link.

But even if the concerns over SGLT2 inhibitors are allayed, many researchers do not think that they are a solution to one of the world’s most debilitating diseases. Kaul wants the FDA to stop approving diabetes drugs just because they lower blood sugar. Instead, he says, companies should prove that their drugs deliver real clinical benefits — reduced complications of the disease, such as kidney damage, amputations and vision loss. He notes that although blood-sugar reduction has been shown to benefit patients who have type 1 diabetes, the same has not been shown for those with type 2 disease.

And for David Nathan, director of the Diabetes Center at the Massachusetts General Hospital in Boston, the problem with SGLT2 inhibitors is not so much their safety as their efficacy. The drugs lower blood sugar by less than half as much as some older drugs. He thinks that the real way to stem the diabetes epidemic might not be in lowering blood sugar, but in addressing excess weight.

But weight-loss drugs have been harder to come by than new diabetes therapies — and they are every bit as fraught with safety concerns. “That’s where the emphasis should be,” says Nathan. “But that’s been a sad story.” ■

SOURCE: IDF

AGRICULTURE

Food fuelled with fungi

Ecologists are starting to appreciate the power of microbes to make crops hardier.

BY NICOLA JONES

With the planet's population booming and climate change threatening traditional 'bread-basket' regions, researchers are seeking ways to squeeze more food from the land. Some are taking a side-ways approach: instead of trying to produce hardier crops through breeding or genetic modification, they are manipulating the vast array of symbiotic microorganisms that live in plants.

Next spring, Adaptive Symbiotic Technologies in Seattle, Washington, will bring to market the first commercial product that harnesses such microorganisms — known as endophytes — to improve crops. The company plans to sell a mixture of fungi for coating rice and maize (corn) seeds, which it says will produce crops with high yields and reduced water use even in harsh conditions. "It's a real paradigm shift in plant ecology," says company founder and plant biologist Rusty Rodriguez. "Up till now we have focused on plants as individuals, as we have with animals."

In the same way that biologists are now starting to understand the power and influence of the trillions of microbes living in and on the human body, ecologists are getting to grips with plant microbiomes. The result is powerful. Instead of having to find and introduce into a crop a single gene for a coveted trait such as salt tolerance, researchers can use a slew of interacting genes that comes pre-integrated in a living organism, such as a symbiotic fungus.

Conventional breeding has helped to create varieties with increased tolerance to drought, but progress in introducing new genes through genetic manipulation has been slow. Despite decades of research, only one drought-tolerant genetically modified crop has been approved in the United States: Monsanto's DroughtGard maize, which expresses a stress-response gene from bacteria.

Although symbiotic plant-microbe relationships — such as those of the nitrogen-fixing bacteria that live in the roots of legumes — have been known for many decades, applied research in this field is relatively recent. Only in the 1970s did researchers realize that a fungus living in symbiosis with tall fescue grass was responsible for making cattle grazing on infected pastures ill. Scientists in New Zealand later discovered that some endophyte-ridden grasses, although poisonous to livestock, were resistant to attack by weevils. This spawned a niche industry that develops and markets

endophyte-hosting turf varieties that repel pest attacks without being toxic to animals.

Now some researchers are applying similar philosophies to food crops. The approach bucks a trend of sterilizing and simplifying crops, says Rodriguez. "Agriculture has spent the past century wiping out the microbes living in our plants, through pesticides and fertilizers. Now we're trying to reverse that."



Maize could survive drought with the help of fungi.

Endophyte researcher James White at Rutgers University in New Brunswick, New Jersey, agrees. "A lot of companies don't think this way — they go for chemical control. They think the microbes get in the way," he says. "It's not the paradigm that these microbes are significantly impacting plants. But they are." There are thought to be millions of endophytic microbes in the world; only a fraction have been identified, and any given plant can host hundreds.

Rodriguez's work began by happy accident. In the early 2000s, while studying the dozen or so plant species that can survive at 50 °C in the hot soils near geothermal vents in Yellowstone National Park in Wyoming, he found that all of them carried a symbiotic fungus. Although neither the plants nor the fungi could tolerate soil temperatures of 40 °C by themselves, together they could (R. S. Redman *et al. Science* 298, 1581; 2002).

Rodriguez and his colleagues later discovered that the fungi were easily transferable: they could grow in anything from watermelons

to maize and confer heat- and drought-tolerance on those crops. "The endophytes somehow protect the plants from oxidation, so the plants don't turn up all their stress defences," says Rodriguez.

Those findings led him to look for other endophytes optimized to tackle the problems likely to be caused by particular food crops by climate change (R. S. Redman *et al. PLoS ONE* 6, e14823; 2011). The result is a commercial mix of about half a dozen fungi that the team named BioEnsure. Field tests done or commissioned by the company show that, compared to untreated seeds, the product increased maize yields by 85% in Michigan during a 2012 drought, increased seed germination rates by two to five times during 5 °C cold snaps, and enabled maize to use one-third less water. In rice, the scientists saw yield increases of 3–6% in 2012 and 2013, despite drought and early-season planting when temperatures were cool. The crop also used 25–50% less water than normal.

BioEnsure has been approved for use by the US Food and Drug Administration and the Department of Agriculture, and independent tests have shown the mixture to be non-toxic. Rodriguez plans eventually to produce targeted endophyte mixes for more crops, including soya beans, wheat, barley and sugar cane.

But the question of whether BioEnsure will work in commercial conditions is hard to answer: although Adaptive Symbiotic Technologies' field-test results are public, they have not been peer reviewed. Richard Richards, who leads research to breed better wheat for the Australian Commonwealth Scientific and Industrial Research Organisation's plant industry division in Canberra, is dubious. "Typically, there is a metabolic cost of hosting an endophyte, so that crops with endophytes are likely to grow less and be less productive," he says. Rodriguez counters that "in all the field work we've done over 15 years we haven't seen anything suggesting metabolic cost".

Others are cautiously optimistic. Mogens Nicolaisen, who works with plant pathogens at Aarhus University in Denmark, thinks that endophytes could be a good way to help introduce resistance to both drought and disease, including pathogens such as wheat rust, an area that Rodriguez says he is pursuing. But, Nicolaisen adds, getting the endophytes into seeds and regulating their growth in different environmental conditions will be tricky. "It will be very hard to control," he says. ■

SAUL LOEB/AFP/GETTY



Mining in the Western Ghats region, where around one-third of India's plant and animal species are found.

CONSERVATION

India faces uphill battle on biodiversity

Government decision to limit protection for species-rich mountains angers conservationists.

BY T. V. PADMA

India's Western Ghats are a tainted paradise. Running almost the length of the country's western coast, the mountain range covers just 6% of India's landmass but is home to more than 30% of its plant, fish, bird and mammal species, making it one of the world's top ten biodiversity hotspots. But the mountains also contain large mineral reserves.

The question of how to strike a balance between protecting and developing the region, home to 39 UNESCO World Heritage sites, has been troubling India. Last month, matters came to a head when ministers announced that they would accept the recommendations of a working group to cordon off more than one-third of the region and ban many industrial activities within it.

This may sound like good news, but the recommendations ran roughshod over a 2011 government-commissioned report by the Western Ghats Ecology Expert Panel. Headed by one of India's leading ecologists, Madhav Gadgil, a visiting professor at Goa University in Taleigao Plateau, the report advised classifying the entire region as ecologically sensitive.

The two groups are now at loggerheads, and the government's November decision has led to protests by conservationists, farmers and the mining and construction industries.

"There was no need for yet another report after the Gadgil committee report," says Sreedhar Ramamurthy, managing trustee of the Environics Trust, a non-governmental organization based in New Delhi. "The aim seems to be to open up more areas for development projects."

Much of the concern over the Western Ghats (see 'Protection limits') centres on the rampant, and highly polluting, illegal mining of iron ore and manganese in Goa and Karnataka states.

In 2010, environment minister Jairam Ramesh commissioned Gadgil to look at the region's ecological and industrial problems. The report, submitted in August 2011, graded

the entire Western Ghats into three zones of varying ecological sensitivity and recommended a ban on mining in the most sensitive areas.

The report came a month too late for Ramesh, who was replaced in July 2011 after rattling the powerful mining lobby by enforcing environmental standards. The whole issue was shelved until August 2012, when Ramesh's successor, Jayanthi Natarajan, commissioned a panel headed by Krishnaswamy Kasturirangan, the former chairman of India's space agency, to look again at the region.

The resulting report recommends designating just 37% of the Western Ghats as ecologically sensitive — some 60,000 square kilometres including World Heritage sites, protected areas and tiger reserves. Banned from this zone would be quarrying and mining activities, as well as other heavily polluting industries such as the manufacture of pesticides and cement.

Almost all of the remaining area (60%) is designated as 'cultural landscape', covering villages, agriculture and non-forest plantations.

The authors acknowledge that they "deviated" from Gadgil's report by recommending prohibitions only on the most damaging activities in the most sensitive areas, and instead placed a focus on "good development". Industry bodies view the Kasturirangan report, which was accepted by the government on 16 November, as an improvement on Gadgil's.

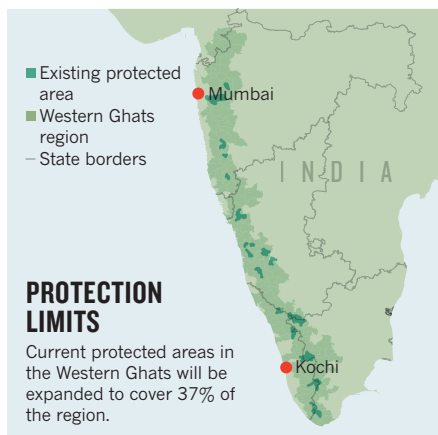
But the 37% demarcation has angered biodiversity experts. They say that the panel has in effect left the rest of the Western Ghats open to industrial activities. On reading a draft of the report in May, Gadgil shot off an open letter to Kasturirangan, saying that the decision is "like trying to maintain oases of diversity in a desert of ecological devastation".

In an interview with *Nature*, Gadgil added, "The report has arbitrarily come up with the concept of 'natural' and 'cultural' landscapes, suggesting that only 'natural' landscapes need to be protected. This has no scientific basis."

Narendra Prasad, former head of the landscape ecology division at the Salim Ali Centre for Ornithology and Natural History in Coimbatore, thinks that the Gadgil report is "more acceptable" given that it favours a "more democratic approach" and long-term sustainability.

Kasturirangan did not respond directly to Gadgil's criticism, but told *Nature* that issues raised by Gadgil "have already been well addressed" in his report. Government officials declined to comment.

Gadgil notes that India's 2002 Biological Diversity Act has sufficient provisions to protect the nature and people of the Western Ghats — if only the conditions were enforced. But "successive governments have done their best to scuttle them", he says. The government's acceptance of the new report is yet another step, he adds, towards "development by imposition, and conservation by imposition". ■



COSMOLOGY

Life possible in the early Universe

Planets orbiting the first stars could have been habitable, challenging arguments for a multiverse.

BY ZEEYA MERALI

Aliens might have existed during the Universe's infancy. A set of calculations suggests that liquid water — a prerequisite for life — could have formed on rocky planets just 15 million years after the Big Bang.

Abraham Loeb, an astrophysicist at Harvard University in Cambridge, Massachusetts, has realized that in the early Universe, the energy required to keep water liquid could have come from the cosmic microwave background, the afterglow of the Big Bang, rather than from host stars. Today, the temperature of this relic radiation is just 2.7 kelvin, but at an age of around 15 million years it would have kept the entire Universe at a balmy 300 kelvin, says Loeb, who posted his calculations to the arXiv preprint server this month (<http://arxiv.org/abs/1312.0613>).

Loeb says that rocky planets could have existed at that time, in pockets of the Universe where matter was exceptionally dense, leading to the formation of massive, short-lived stars that would have enriched these pockets in the heavier elements needed to make planets. He suggests that there would have been a habitable epoch of 2 million or 3 million years during which all rocky planets would have been able to maintain liquid water, regardless of their distance from a star. “The whole Universe was once an incubator for life,” he says.

Loeb's result also challenges the anthropic principle, a line of reasoning that is invoked to explain why certain physical parameters seem to be tuned to the precise values needed for life: the Universe is the way it is because beings exist to observe it. The principle is consistent with the idea of a multiverse: if multiple universes exist, each based on different parameters, then intelligent beings should not be surprised to find themselves in one in which those parameters are suited to life.

In the 1980s, Nobel laureate and physicist Steven Weinberg used an anthropic argument to calculate a maximum value for a measure

of the intrinsic energy of the vacuum in space that, in theory, would push space outwards. Weinberg pointed out that unless this value is tiny, it would have torn matter apart before the Sun, Earth or humans could have come into existence. His prediction seemed to be confirmed in the late 1990s, when astronomers discovered dark energy, which seems to act like a vacuum force that accelerates the expansion of the Universe — but by only a small amount. According to particle physicists' calculations, dark energy should actually be some 120 orders of magnitude stronger than Weinberg's maximum value. Multiverse proponents take this as evidence that multiple universes exist that have higher values of vacuum energy.

But during Loeb's proposed habitable epoch, matter was so dense that even if the vacuum energy had been a million times stronger, it would not have prevented the formation of stars and rocky planets, and the emergence of life. Thus, Loeb says, advocates of the anthropic principle cannot claim that the small value observed now is the only one that could be observed by living beings.

Responses to Loeb's work vary. Christopher Jarzynski, a biophysicist at the University of Maryland, College Park, is not convinced that life could exist in a uniformly warm Universe. Life on Earth depends thermodynamically not only on the heat source of the Sun, but also on the cold cosmic microwave background, which provides a heat sink, he notes. “Life feeds off this,” he says. And Alexander Vilenkin, a cosmologist at Tufts University in Medford, Massachusetts, says that a few million years is too short a time to produce intelligent life.

Yet Freeman Dyson, a physicist at the Institute for Advanced Study in Princeton, New Jersey, thinks that life might be more adaptable than we think. “Anything is habitable if you are clever enough,” he says. ■

“The whole Universe was once an incubator for life.”

CORRECTION

The News story ‘Image search triggers Italian police probe’ (*Nature* **504**, 18; 2013) should have said that paper identified as problematic and being retracted was in *Cell Death and Differentiation*, not *Cell Death and Disease*.



THE MEDICAL TESTING CRISIS

With a serious shortage of medical isotopes looming, innovative companies are exploring ways to make them without nuclear reactors.

BY RICHARD VAN NOORDEN

In 2009, two nuclear research reactors shut down for repairs and maintenance. This was not surprising, given that both were around half a century old. But these reactors happened to produce most of the world's supply of the radioactive tracer technetium-99m, an isotope injected into patients in 70,000 diagnostic scans a day. Hospitals around the world went into a panic.

Finding themselves suddenly short of the crucial isotope, doctors cancelled scans, postponed operations or switched to older diagnostic techniques that exposed patients to higher doses of radiation. "It was the isotope equivalent of an electricity blackout," says Ronald Schram, who manages one of the affected reactors, the High Flux Reactor at Petten in the Netherlands. Nobody knows exactly how much damage was done, says Fred Verzijlbergen, head of the department of nuclear medicine at Erasmus Medical Center in Rotterdam, but "it was very serious. Many hospitals didn't receive technetium for weeks."

The crash made it painfully clear that the world's medical-isotope supply chain was dangerously fragile, relying heavily on about four government-subsidized reactors built in the 1950s and 1960s. Isotope supplies have taken a hit again and again, most recently last month,

when Canada's Chalk River reactor shut down unexpectedly for a few days at the same time as two other reactors. And more shortages are coming. The Chalk River reactor, which produces close to one-third of current global supplies, is slated to end production of isotopes in 2016.

But for nuclear engineer Greg Piefer, the crisis presents an opportunity. In 2005, fresh out of a nuclear-engineering doctorate at the University of Wisconsin–Madison, he had dreamed up a way to use particle accelerators — rather than nuclear reactors, with their problematic waste — to transform uranium into technetium. His idea did not get much attention at the time. After the 2009 disaster, however, politicians demanded new ways of making medical isotopes, particularly in the United States, which accounts for 50% of world medical-isotope demand but has no local production capacity. Piefer's ideas, and those of other aspiring entrepreneurs, were thrown into the spotlight.

TECHNICAL CHALLENGE

At least five North American companies and collaborations, including Piefer's firm, SHINE Medical Technologies in Madison, are pioneering methods that should produce medical isotopes in the next few years. It

Many diagnostic scans rely on radioactive technetium-99m.

is not clear which will win out — nor whether they can replace the conventional reactor approach or be ready soon enough to avert another shortage. “It’s boiling down to a rather critical situation in 2015–16,” says Schram.

Technetium-99m is often called the workhorse of modern medical imaging, because it accounts for about 80% of the world’s use of radioactive isotopes in nuclear medicine, 90% of which is diagnostic scans. A γ -ray emitter with a half-life of just 6 hours, it can be attached to a molecule that targets the organ of interest. Medical scans known as single-photon emission computed tomography (SPECT) then pick up the glow of the radioisotope. Such tests are used to check how well blood is flowing to heart muscles, to spot whether cancers have spread through bones and to assess blood flow in the brain.

The creation of technetium-99m involves one of those miraculous, globe-crossing supply chains that modern economies have rendered commonplace. The journey starts with enriched uranium from the United States, which is made into plates and shipped to research reactors around the world. Each plate is baked for a week in the glare of a nuclear reactor’s neutrons, which fission about 6% of the uranium into molybdenum-99. This has a half-life of 66 hours and slowly decays into technetium. Hospitals across the globe purchase ‘moly cows’ — paint-tin-sized devices that hold the molybdenum-99 bound tightly to alumina. By flushing the technetium out with saline solution, hospital technicians can milk the moly cows for fresh supplies of technetium for up to two weeks.

Nuclear reactors are the most efficient way to produce molybdenum, says Benard Ponsard, who manages the isotope-producing BR2 reactor in Mol, Belgium. Many others agree. In the 1990s, Canada planned new reactors that would have circumvented many of the shortages, but these were mothballed in 2008 after technical problems emerged that proved too expensive to fix. So other countries are now racing to fill the gap. BR2 aims to start upgrading its medical-isotope capacity at the end of 2014. The OPAL reactor in South Sydney, Australia, is planning upgrades that would quadruple its isotope supplies by 2017. New reactors or upgrades are planned from Argentina to China (see ‘Supply fix’).

This promises a lot of production capacity. But there could still be problems. The new reactors might not supply enough isotopes to provide a sufficient cushion in case of major breakdowns, notes Robert Atcher, director of the US National Isotope Development Center, created in 2009 by the Department of Energy to help to manage isotope distribution.

PRICE HIKES

More crucially, the cost of reactor-sourced molybdenum could skyrocket. Because the reactors are involved in research, they are subsidized by their host governments and sell their molybdenum at below-market prices. This means there is little incentive for companies to invest in new production facilities, concluded a post-crisis review by the Nuclear Energy Agency (NEA) of the Organisation for Economic Co-operation and Development (OECD) in Paris. The NEA has endorsed a plan to end the subsidies, which those in the supply chain are now preparing for. According to Ron Cameron, head of nuclear development for the agency, the price of molybdenum from reactors could increase as much as sevenfold when that happens. Meanwhile, the United States has decided to stop exporting highly enriched uranium, because it might be intercepted to make nuclear weapons. By 2020, reactors will have to make do with low-enriched uranium fuel and plates — which could increase molybdenum costs from reactors by another 40%. The final price to the middle-man is “very much up in the air,” says Atcher, who guesses it might increase 15-fold.

That has led some to seek a radical alternative. Rather than depending on a few distant centres of production, each costing hundreds of millions of dollars, hospitals could get their isotopes locally from facilities that

have small medical cyclotrons costing just a few million dollars, says Paul Schaffer, head of the nuclear medicine division at TRIUMF, Canada’s national laboratory for particle and nuclear physics in Vancouver, and leader of a team pursuing this idea.

In the cyclotron model, neither nuclear reactors nor uranium are needed. A beam of accelerated protons shoots into a target of molybdenum-100, creating technetium-99m directly. Technetium’s 6-hour half-life means that the product cannot be transported far: a single cyclotron could cover maybe a 400-kilometre radius. The idea is to have lots of cyclotrons distributed across major urban areas. That is not as ambitious as it might seem, points out TRIUMF spokesman Tim Meyer: many hospitals already use in-house cyclotrons to produce isotopes for a more advanced form of imaging, positron emission tomography.

In June, the TRIUMF team announced that by running an upgraded cyclotron in Vancouver overnight, it could make enough technetium to satisfy the city’s needs. “The dozen or so cyclotrons [already] in Canada could cover 90% of the population and 50% of the geography” when adapted with a TRIUMF upgrade kit, says Meyer. The team awaits approval from Health Canada, which will confirm that its technetium is safe for use. Schaffer says that cyclotrons will not produce enough technetium to supply all of Canada’s needs by 2016. “But a decentralized supply is certainly possible in the long term,” he says.

Other countries seem to be interested in the strategy. Advanced Cyclotron Systems — a firm in Richmond, Canada, that sells cyclotrons and is working to make cyclotron technetium — has had inquiries from the United Kingdom, Saudi Arabia, Thailand and more. “It’s got health authorities in many countries quite excited,” says John Taylor-Wilson, the company’s vice-president of marketing and business development.

Atcher, however, is not convinced that the approach will be helpful in the United States, where, by chance, most hospitals have lower-power cyclotrons that could not produce as much technetium. And if a cyclotron goes down for repair, an urban area might be left without a back-up plan.

BRIGHT IDEAS

Piefer has a plan that is a little less radical than the cyclotron model. His team at SHINE wants to stick with the current distribution system, but get rid of the expensive nuclear reactor at the system’s heart.

SHINE’s technology uses a linear accelerator to slam deuterium ions into tritium gas, producing helium and neutrons. The neutron flux is orders of magnitude less than that emerging from a nuclear reactor. Instead of a small uranium plate, the neutrons are fired into a couple of hundred litres of warm, low-enriched uranium salts. The molybdenum can be rinsed away using ion-exchange resins, and the unconverted uranium recycled for use in the same facility. Nuclear waste is produced from the uranium target, but it is only a small fraction of that produced from reactor fuel, notes Piefer.

Piefer hopes to build a facility in Janesville, Wisconsin, which he says would supply half of the United States’ need for technetium (about one-quarter of world demand). But construction has not started yet, because the firm — like any US company working with uranium — has to wait for a permit from the Nuclear Regulatory Commission. And it needs to raise another US\$150 million, on top of \$30 million raised so far from a Department of Energy grant and investors. “If sufficient funding were available, we’d start production by the end of 2016,” Piefer says — about the same time the Chalk River reactor ends its supply.

Piefer will face some competition from right across town. Madison is home to another innovative medical-isotope company: NorthStar, which also hopes to supply half the United States’ medical isotope needs in the next few years — and in the

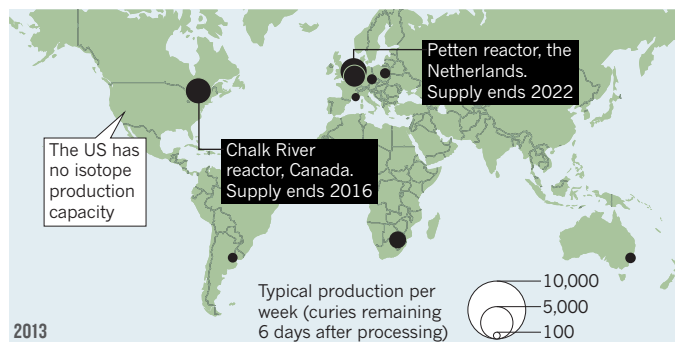
“IT’S BOILING DOWN TO A RATHER CRITICAL SITUATION IN 2015–16.”

➔ NATURE.COM

For a podcast about alternative isotope technologies, see: go.nature.com/w74z5a

SUPPLY FIX

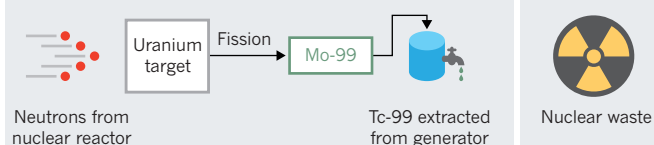
Today, just a few nuclear reactors provide global supplies of technetium-99m (Tc-99), the most popular tracer used in medical scans. Many of the reactors are old and some will soon stop producing medical isotopes, leaving the world vulnerable to shortages.



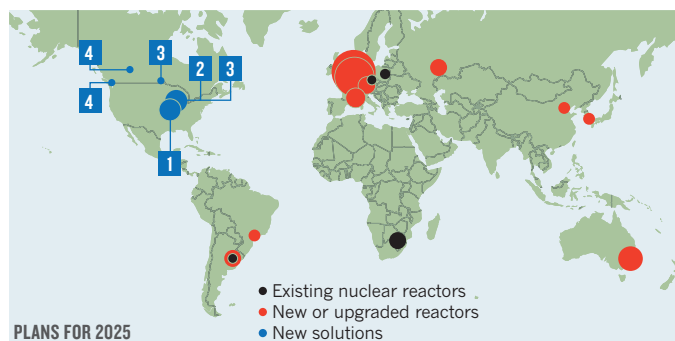
PROCESS

BY-PRODUCT

NUCLEAR REACTOR

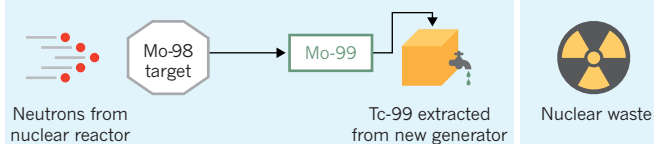


New nuclear reactors will diversify isotope supplies in coming years. But researchers are also developing production methods that avoid using a reactor or a uranium target. Some schemes produce little or no nuclear waste.

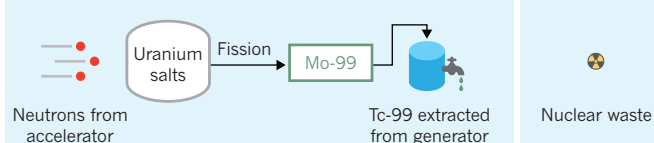


PLANS FOR 2025

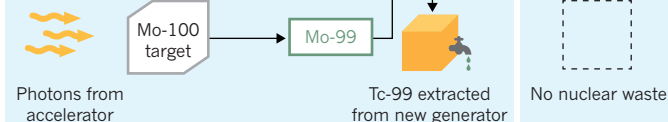
1 REACTOR WITH NEW TARGET (NORTHSTAR)



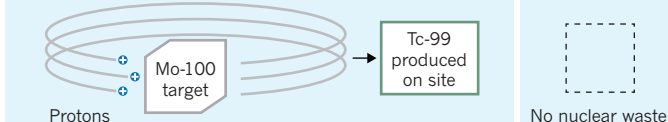
2 ACCELERATOR (SHINE MEDICAL TECHNOLOGIES)



3 ACCELERATOR (NORTHSTAR/PRAIRIE ISOTOPE PRODUCTION ENTERPRISES)



4 CYCLOTRON (TRIUMF/ADVANCED CYCLOTRON SYSTEMS)



longer term, aims to double that output with a second project. The fact that it has ended up so close to SHINE is “a bizarre coincidence”, says Piefer; the rivals are “cordial but competitive”, he adds.

Whereas SHINE avoids using a nuclear reactor but keeps a uranium target, NorthStar aims to do the opposite. It ditches the uranium target but — at least at first — still makes use of a reactor. The company’s short-term plan is to use the research nuclear reactor at the University of Missouri in Columbia to fire neutrons into a molybdenum-98 target, making molybdenum-99. This would be very quick to get up and running, but it invites complications on another front. Rather than being able to separate molybdenum-99 from the surrounding uranium, which is relatively easy, the company has the trickier task of separating the desired isotope from the molybdenum starting material. So NorthStar has had to design a generator to replace the moly cow. The result, about the size of a microwave oven, requires a computer, pipes and valves to extract technetium — more complex than the simple saline wash that hospitals use today.

In March, NorthStar submitted a New Drug Application to the US Food and Drug Administration (FDA), seeking to prove that its milked technetium is equivalent to what comes out of moly cow generators. It is hoping to receive the final word by the end of the year, says James Harvey, the company’s chief science officer. “The Missouri project will be fully in production by mid-2014,” Harvey says.

In the long-term, NorthStar has a more ambitious plan that would again cut reactors out of the system. It plans to use high-energy photons from a linear accelerator to kick neutrons out of molybdenum-100 to produce molybdenum-99. (Prairie Isotope Production Enterprises, a non-profit firm based in Winnipeg, Canada, is looking at a similar system, but on a smaller scale.) Again, NorthStar would need to use its microwave-sized hospital generator, assuming that it gets FDA approval. And although the company has raised \$50 million already, it will need much more to get this scheme off the ground: Harvey will not reveal a figure, but says it is a lot less than \$150 million.

Atcher has doubts about both NorthStar and SHINE. “In a nutshell, both of these companies are start-up companies,” he says, adding, “2016 is not that far away and they are scrambling to get their programmes going.”

MARKET FORCES

Underlying all the jostling are questions to which no one has good answers: how expensive the technetium from these new technologies will be, and whether the schemes will create enough to replace the nuclear-reactor approach. “Obviously, each of these competitors has a secret recipe where they think they can beat the odds,” says Meyer. But the uncertainty in the economics scared away larger companies such as General Electric and Babcock and Wilcox, both of which initially showed an interest in developing medical-isotope schemes but backed out last year.

Atcher thinks that reactors will always come out ahead. Others see a more diverse future. “The long-term scenario will really be driven by the market,” says Schaffer. “I equate it to the electricity market, where we have nuclear, wind, hydroelectric, solar and so on. And the price of that source of electricity pretty much defines its share of the market. I believe the same thing will happen with isotopes, with sources from cyclotrons, linear accelerators and nuclear reactors.”

So can hospitals avert catastrophe in 2016, when the Chalk River production facility shuts down? “On paper, it looks like the world can compensate,” says Schaffer. But, he adds “it’s such a dynamic situation”. Cameron, more soberly, says that the answer will not be clear until he sees which reactors and companies ramp up production. “We have to do a lot of sums to see how the effects will balance out.”

All the uncertainty about the technologies is leaving doctors such as Verzijlbergen concerned. “There is a lot of optimism but we need proof,” he says. “From the medical side I am interested in reliable supply.” He still worries that in a few years he will have to go to patients with difficult news: that a widespread shortage of technetium means that they cannot have the diagnostic tests that they need. ■

Richard Van Noorden is a senior reporter at Nature in London.



UNDER THE VOLCANO

GEOPHYSICISTS ARE SCOURING THE GLOBE FOR EVIDENCE OF MANTLE PLUMES — THE PRESUMED SOURCE OF SOME MEGA-ERUPTIONS.

BY ALEXANDRA WITZE

Karin Sigloch sailed into the Indian Ocean this autumn to retrieve traps she had set out the year before. She wasn't hunting deep-sea creatures but something deeper and more elusive — a thin stream of hot rock rising through the bowels of the planet.

Sigloch, a geophysicist at Ludwig Maximilians University Munich in Germany, is one of the latest scientists to pursue this long-sought quarry, known as a mantle plume. Such features are thought to feed some of the most active volcanoes on Earth, and could fuel the biggest volcanic outpourings ever seen. The one Sigloch is tracking, for example, has been linked to eruptions in India 65 million years ago that were so massive they may have contributed to the demise of the dinosaurs.

Mantle plumes could also explain why the Hawaiian volcanoes and many others appear in the middle of crustal plates, far from where they might be expected. And, by releasing pent-up heat from Earth's interior, plumes could have played a major part in how the planet has evolved over billions of years.

The study of mantle plumes has been a mainstay of geophysics since the idea was introduced in the early 1970s¹. Using seismic waves to image the innards of the planet, researchers have captured hints of structures that appear to be plumes. But questions linger over what these deep features really are, and a small but vocal subset of critics continues to question whether plumes exist at all.

This is why Sigloch and her colleague Guilhem Barruol of the University of La Réunion, on the French island of Réunion in the Indian Ocean, have spent the past six weeks trawling the ocean's depths. Battling bad weather, equipment glitches and the occasional shark attack on their instruments, the French-German team hauled up all 57 seismometers that they had left on the ocean bottom in autumn 2012. The instruments have been recording the seismic waves that race from distant earthquakes and shake the sea floor. By analysing those vibrations, the researchers plan to map out the rocks beneath Réunion to see whether a mantle plume feeds the island's main volcano².

If they succeed, they may chart a mantle plume in greater detail than has ever been done before. They hope to answer some outstanding questions, such as where plumes originate, whether they rise vertically or corkscrew off to one side, and whether they branch into multiple smaller conduits just before breaching Earth's surface.

"Up until now, people have tended to speak of plumes without direct observation," says Barruol. "Our experiment will be able to see if there is something there."

Ask a geophysicist how many mantle plumes exist, and the answer may range from zero to several dozen. Most scientists, however, settle on 10 to 20 volcanic hot spots, with some of the strongest examples in Hawaii, Réunion and Tristan da Cunha in the South Atlantic Ocean. The problem is knowing exactly how these regions are fed by plumes.

Earthquakes offer a tool for answering that question. Because seismic waves travel more slowly through

The Piton de la Fournaise volcano on the French island of Réunion may sit atop a mantle plume.

molten or soft rock than through solid stone, researchers can map out temperature and density differences inside Earth by analysing how quickly seismic waves arrive at different stations after an earthquake. This allows them to probe the planet's outer crust and the underlying mantle — all the way down to the boundary with the core, some 2,900 kilometres below the surface.

The problem is that seismic images struggle to pick up narrow features, and plumes are thought to be just several hundred kilometres across until they hit the underside of the planet's outer shell, or lithosphere, where they spread out to form the plume head (see 'Deep heat'). "They should be there, but it's hard to see them," says Sigloch, who is moving to the University of Oxford, UK, in January.

The key to getting better seismic images is to blanket Earth's surface with instruments. On land, this approach has worked well at places such as Yellowstone National Park in Wyoming, thought to lie atop a mantle plume responsible for some of the biggest volcanic blasts in the past 2 million years. Those data were gleaned from a project called EarthScope that has been creeping across the contiguous United States for the past nine years (see *Nature* **503**, 16–17; 2013), deploying temporary seismometers in a 70-kilometre-square grid. It allowed seismologists to generate the best picture yet of the Yellowstone plume³, revealing the structure to be a hot, narrow upwelling that reaches into the lower mantle, at least 900 kilometres deep.

But many mantle plumes are thought to

underlie the oceans — and putting instruments on the sea floor is more difficult and expensive than deploying them on land. In the early 2000s, Barruol took a first stab at studying a presumed oceanic plume by installing ten seismic stations on islands in French Polynesia. That turned out to be too few to determine whether a mantle plume was rising beneath the ocean floor in the South Pacific.

In 2005, a University of Hawaii-led team upped the game, placing 36 ocean-bottom seismometers around the Hawaiian islands for a year and then shuffling them to different positions for a second year. The project, which went by the acronym PLUME, yielded the best images yet of some kind of deep structure extending far beneath Hawaii⁴.

WHAT LIES BENEATH

But the more researchers looked at the Hawaiian plume, the more complicated things got. In the classical plume paradigm, a jet of hot, viscous rock rises to the bottom of the lithosphere, where it pools in a sort of pancake some 100 kilometres thick before trickling upwards to feed volcanoes. But under Hawaii, the seismic images suggest a different structure, with an unexpected bulge in the plume well below the bottom of the lithosphere.

Maxim Ballmer, a geophysicist at the University of Hawaii at Manoa, thinks rock chemistry can be used to explain what is going on⁵. If the plume is rich in a mineral called eclogite, which is denser than typical mantle materials, it will stall as it reaches a depth of about 400 kilometres. The plume rock should pool there and spread out horizontally. Eventually, after more heat rises from below, the eclogite-rich rock would become buoyant enough to rise as a thin plume, Ballmer says.

Researchers have also found other oddities beneath Hawaii. A team led by Catherine Rychert at the University of Southampton, UK, used data from the PLUME experiment to detect a warm pool about 110–155 kilometres below the surface, but it seems to be centred about 100 kilometres west of the archipelago's main island rather than directly beneath it⁶. That could mean that the Hawaiian plume bends as it approaches the surface, Rychert says — maybe because of some kind of chemical boundary that diverts the plume's flow to one side, or perhaps because the plume rises along a skewed path to begin with.

Beyond that, there is some dispute about how deep beneath Hawaii the PLUME experiment could reliably measure — perhaps only to 1,500 kilometres or so. One of Sigloch and Barruol's goals is to use the dense array of instruments around Réunion to hunt even deeper for its plume. "We are hoping to complete this picture into the lower half of the mantle, which wasn't possible in the case of Hawaii," says Sigloch.

The problem with Hawaii is that it lies in the middle of the Pacific Ocean. Réunion

has the advantage of being close to Madagascar and relatively close to southern Africa, where researchers can place seismic instruments more easily. Sigloch and Barruol are using data from Madagascar and Africa, along with some islands in the Mozambique Channel, to supplement the ocean-bottom seismometers they deployed last year. Altogether, about 120 seismic stations are being used for the Réunion study, covering an area some 2,000 by 3,000 kilometres.

But even with all that, the Réunion work may not be definitive, says Cecily Wolfe, a seismologist at the University of Hawaii who led the PLUME study. "It really depends on what the structure is and what techniques and tricks they can apply to image it," she says.

No matter what the Réunion study finds, its results are unlikely to convince the few critics of the plume hypothesis. One of them is Warren Hamilton, a geophysicist at the Colorado School of Mines in Golden. He says that 'plumologists' are biased towards seeing plumes in seismic data.

There's no question that plumes are more complicated than once thought. At this week's American Geophysical Union meeting in San Francisco, California, Rychert is scheduled to talk about her seismic studies of several possible plumes in areas including Hawaii, the Galapagos islands, Iceland and the Afar region of eastern Africa. The particular technique she uses probes only the upper few hundred kilometres and so cannot see deep into the mantle. But in most cases, as in Hawaii, she sees features about 100 kilometres deep that she thinks could be ponds of molten rock fed by plumes. And they aren't always where scientists might expect.

In the Galapagos, Rychert has pinpointed three such features: one right where a mantle plume has been identified in the past and two in other locations. That could mean that plumes are smearing around as they hit the lithosphere, diverting and feeding separate volcanoes. "These seismic methods allow us to look in greater detail, and sometimes we see more complicated phenomena," she says. "Those are the exciting parts."

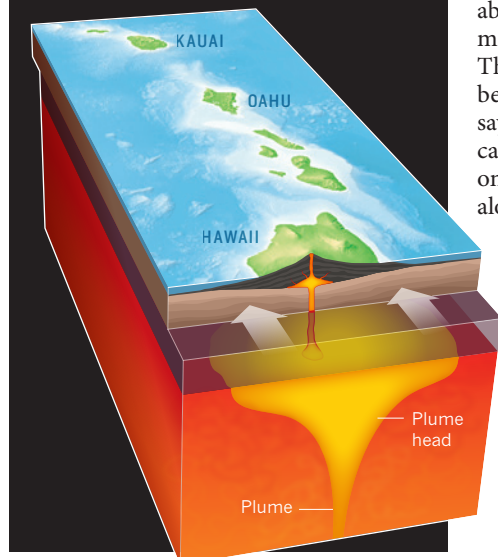
Sigloch and Barruol are expecting plenty of excitement in the coming months. Last week, they pulled into port at Réunion, their ship stacked with seismometers and precious data. It will be at least a year before they crunch through all that information and can see what is happening beneath Réunion. But the underwater traps seem to have caught their prey. ■

Alexandra Witze writes for *Nature* from Boulder, Colorado.

1. Morgan, W. J. *Nature* **230**, 42–43 (1971).
2. Barruol, G. & Sigloch, K. *Eos* **94**, 205–207 (2013).
3. Schmandt, B., Dueker, K., Humphreys, E. & Hansen, S. *Earth Planet. Sci. Lett.* **331–332**, 224–236 (2012).
4. Wolfe, C. J. *et al. Science* **326**, 1388–1390 (2009).
5. Ballmer, M. D., Ito, G., Wolfe, C. J. & Solomon, S. C. *Earth Planet. Sci. Lett.* **376**, 155–164 (2013).
6. Rychert, C. A., Laske, G., Harmon, N. & Shearer, P. M. *Nature Geosci.* **6**, 657–660 (2013).

DEEP HEAT

A plume of hot rock rising beneath Hawaii is thought to fuel the island's active volcanoes. Movement of the oceanic plate over that plume has left a chain of older volcanoes such as those on the islands of Kauai and Oahu.



COMMENT

BIBLIOMETRICS Women's research output and citations analysed **p.211**

ANIMATION Gollum co-creator on digital storytelling **p.214**



DOCUMENTARY Celebrating John Milne, founder of modern seismology **p.216**

AGRICULTURE Funds needed to fight banana fungus ravaging southeast Asia **p.218**

DAVID PARKINS



Support mothers to secure future public health

Evidence that long-term health is shaped by the environment in early life calls for prenatal interventions to tackle chronic disease, argue **David Barker** and colleagues.

Worldwide, chronic diseases are on the rise. Globally, the prevalence of adult-onset diabetes is expected¹ to approximately double by 2030, and that of cardiovascular disease to increase by about 35% (ref. 2). Current approaches to curbing this situation are failing.

More than 30 years of epidemiological studies using data from several hundred thousand people around the world point to factors during prenatal and early childhood development that contribute to these statistics.

Together, these data suggest that instead of focusing exclusively on people's genes, or

on their diets and lifestyles in adulthood, we need a developmental approach to public health. People working in public health must support girls and young women with low incomes to help them to feel more in control of their lives and so better able to prioritize healthy eating. At the same time, better access to quality food is necessary so that it is easier for people to make healthier choices. This would improve mothers' nutrition and thereby the health of future generations.

Clues that chronic disease might result from perturbations in growth during early life came from an unusual data source.

Nearly 30 years ago, an exhaustive search of every record office in every town and city in England and Wales unearthed ledgers in the county archive of Hertfordshire containing health visitors' records for thousands of babies born between 1911 and 1948. These records documented the babies' weight at birth and at one year old, whether they were fed breast milk or formula, the age at which they were weaned, and significant health events, such as bouts of illness, in their early life.

Tracing many of the individuals named in these records revealed that their health in adulthood was associated with how they ▶

► had grown in the womb. The Hertfordshire data and similar records from other UK towns revealed, for instance, that a person weighing 2.7 kilograms (6 pounds) at birth has a 25% higher risk of contracting heart disease in later life, and a 30% higher risk of having a stroke, compared with someone weighing 4.1 kilograms (9 pounds) at birth³.

These findings were soon strengthened by data from a cohort of 20,000 people born in Helsinki between 1924 and 1944. This study showed, for example, that if all the babies at birth had had weights within the highest third of the total range, the incidence of diabetes in later life would have been halved⁴. In the years since, numerous other studies, involving people from places as diverse as Europe, India, Guatemala, the Philippines and South Africa, have revealed similar correlations with effects that extend to the health of grandchildren.

DEVELOPMENT AND DISEASE

In the past 15 years, researchers have begun to understand the biology underlying the links between development and chronic disease. The evidence suggests that women should start eating healthily well before they get pregnant. Women who are obese, for example, accumulate more metabolites (such as insulin, lactate and triglycerides) in their ovarian follicles⁵ than do women who are not obese. This accumulation can reduce their fertility and increase the likelihood that their offspring will develop certain diseases, such as diabetes, cardiovascular disease or cancer, later in life.

At the moment of conception, the growing embryo seems to be exquisitely sensitive to its nutritional environment. Studies of babies born through *in vitro* fertilization, for instance, have shown that birth weights can be affected simply by changing the constituents of the medium in which the embryos are cultured.

After conception, if resources are in short supply, investment in low-priority organs, such as the kidneys and lungs, which do not function in the womb, may be traded off to protect more important ones such as the brain. Studies in rats and piglets, as well as research using ultrasound to track the growth of organs in human fetuses, show that adverse conditions in the womb are associated with smaller kidneys with fewer nephrons, which regulate the concentration of water and soluble substances in the blood. Fewer nephrons means more wear and tear to existing nephrons, more nephron loss and an increase in blood pressure with age.

A mother's stored nutrients and the turnover of protein and fat in her tissues — a reflection of her lifetime nutrition — seem to be particularly important in later fetal life. Studies of mice, rats and pregnant women have shown that mothers with a greater



The effects of healthy eating before and during pregnancy are far-reaching.

muscle-to-fat ratio and larger organs have a higher rate of protein turnover. This enables them to make amino acids more readily available to their fetuses, especially during the later weeks of pregnancy, when nutritional demands are greatest.

Differences in the size and shape of the placenta may also reflect a mother's nutritional status and serve as biomarkers of a fetus's early experience. These observations come mainly from work on animals, but also from cohort studies relating measurements of mothers' placentas at the time of birth to their diets and the birth weights of their babies. Individuals, who as fetuses were nourished by placentas that are long for their breadth, for example, are up to twice as likely to develop colorectal cancer and, if a first child, up to three times as likely to develop coronary heart disease compared with those who had been nourished by placentas of average size and shape⁶.

And although babies deprived *in utero* but then born into conditions of plenty often 'catch up' in body weight in their first year of life, such compensatory growth has costs. In fish and birds, offspring who start small and grow fast have fewer offspring, are more susceptible to infections and disease, and live shorter lives than offspring that start life larger. The Helsinki data similarly indicate that boys who catch up between birth and age seven do not live as long as those whose growth follows a more standard trajectory.

ACHIEVING CHANGE

Today, millions of fetuses are receiving inadequate nutrition because their mothers are starving, malnourished or overweight. To understand the factors driving poor nutrition, we have been running focus groups and surveys with mothers for the past decade — first, in Southampton, UK,

beginning in 2004, and then in Klamath Falls, Oregon, starting in 2007.

By asking questions to try to understand what healthy eating means to mothers, and how they would ideally like to feed their children, we have established that young women know that they and their children should eat a balanced and varied diet⁷. But other challenges, such as having to balance the cost of food with the potential for waste, leave mothers feeling out of control and with little motivation for the battle involved in creating good eating habits in their families.

Under these circumstances, the public-health approach currently used across industrialized nations, of providing women with information about healthy eating, seems unlikely to be effective. Our work and that of others strongly suggests that a better strategy would be to support women to identify the barriers they face and, where possible, to empower them to generate their own solutions.

Ten years ago, a team of local and international health-care researchers worked with women from villages in Nepal, aiming to reduce the high levels of mortality among them and their newborn babies⁸. Women were encouraged to join discussion groups run by a facilitator who helped them to formulate simple strategies to improve perinatal care that would work in their communities — such as the production and distribution of basic, home-birth delivery kits. Over the following two years, the 8% of the 28,000 village women who attended the meetings spread the messages to others and, solely because of this intervention, infant mortality fell by 30% and maternal mortality by 80%.

A similar empowerment mechanism underpins the 'disease-self-management' programmes that are increasingly being used all over the world, often with striking effects⁹.

In these, people with chronic diseases are brought together to support each other to better manage their conditions.

Since 2009, we have been involved in the Southampton Initiative for Health, which uses an empowering, problem-solving approach to improve the diets and physical activity levels of Southampton's most disadvantaged young women and their children. The programme has involved training the staff of the city's Sure Start Children's Centres — providers of services such as baby clinics, breastfeeding and weaning support, dentistry, parenting and cookery classes — in having conversations that encourage women to identify problems and generate solutions to change behaviour.

Although the data suggest that attending centres staffed by workers using this approach enhances women's sense of empowerment, to improve their nutritional status we need both to help women to feel more in control of their food choices and to make it easier for them to make better choices.

On a small scale, such a multilevel approach has proved effective. Trials in Canada, Australia and the United States demonstrate that the diets of small-town residents can be improved when efforts to enhance people's sense of empowerment in relation to healthy eating are pursued alongside local media campaigns to promote the benefits of eating well, together with programmes that help people to gain better access to fruits and vegetables and skills in food preparation. The challenge is to scale up such efforts to the wider public-health arena, because this means engaging political and commercial interests,

including those of powerful food companies.

We believe that the methods used by people working in public health to engage politicians and food companies need to undergo a similar transformation to those being used to engage individuals. So far,

“We need to help women to feel more in control of their food choices.”

public-health advocates have called for regulation and legislation as a means to improve diets — an increased tax on fatty and sugary foods, for instance. Yet this is unlikely to happen because raising the tax on soft drinks, say, is not in the interests of industry, or of politicians, who are sensitive to industry pressures and to a public desire for cheap soft drinks.

Instead of wagging fingers, we need to generate consensus. Empowering consumers to call for better access to better food will put pressure on politicians to respond to voters, and on the food industry to please their customers.

More than 20 years ago, one of us (D.B.) wrote¹⁰ in this journal that “if more was known about the processes by which the environment in early life influences adult health ... the rise in incidence of ‘Western’ disease [might be] minimized.” Today, we have the knowledge to readily prevent chronic diseases, had we but the will to do so. ■

David Barker died on 27 August 2013, after writing the first draft of this Comment (see *Nature* **502**, 304; 2013). Before

his death, he was professor of clinical epidemiology at the Medical Research Council (MRC) Lifecourse Epidemiology Unit, University of Southampton, UK; professor of cardiovascular medicine at the Heart Research Center, Oregon Health and Science University in Portland; and visiting professor at the Center for the Study of Human Health, Emory University, Atlanta, Georgia. **Mary Barker** is senior lecturer in psychology at the MRC Lifecourse Epidemiology Unit, University of Southampton, UK. **Tom Fleming** is professor of developmental biology at the Centre for Biological Sciences, University of Southampton, UK. **Michelle Lampl** is director of the Center for the Study of Human Health and Professor of Anthropology at Emory University, Atlanta, Georgia.
e-mail: meb@mrc.soton.ac.uk

1. Chen, L., Magliano, D. J. & Zimmet, P. Z. *Nature Rev. Endocrinol.* **8**, 228–236 (2012).
2. Laslett, L. J. et al. *J. Am. Coll. Cardiol.* **60**, S1–S49 (2012).
3. Barker, D. J. P. *Brit. Med. Bull.* **53**, 96–108 (1997).
4. Barker, D. J. P., Eriksson, J. G., Forsén, T. & Osmond, C. *Int. J. Epidemiol.* **31**, 1235–1239 (2002).
5. Robker, R. L. et al. *J. Clin. Endocrinol. Metab.* **94**, 1533–1540 (2009).
6. Barker, D. J. P. & Thornburg, K. L. *Placenta* **34**, 841–845 (2013).
7. Barker, M. et al. *Pub. Health Nutr.* **11**, 1229–1237 (2008).
8. Manandhar, D. S. et al. *Lancet* **364**, 970–979 (2004).
9. Bodenheimer, T., Lorig, K., Holman, H. & Grumbach, K. J. *Am. Med. Assoc.* **288**, 2469–2475 (2002).
10. Barker, D. J. P. *Nature* **338**, 371–372 (1989).

Global gender disparities in science

Cassidy R. Sugimoto and colleagues present a bibliometric analysis confirming that gender imbalances persist in research output worldwide.

Despite many good intentions and initiatives, gender inequality is still rife in science. Although there are more female than male undergraduate and graduate students in many countries¹, there are relatively few female full professors, and gender inequalities in hiring², earnings³, funding⁴, satisfaction⁵ and patenting⁶ persist.

One focus of previous research has been the ‘productivity puzzle’. Men publish more papers, on average, than women⁷, although the gap differs between fields and subfields. Women publish significantly fewer papers

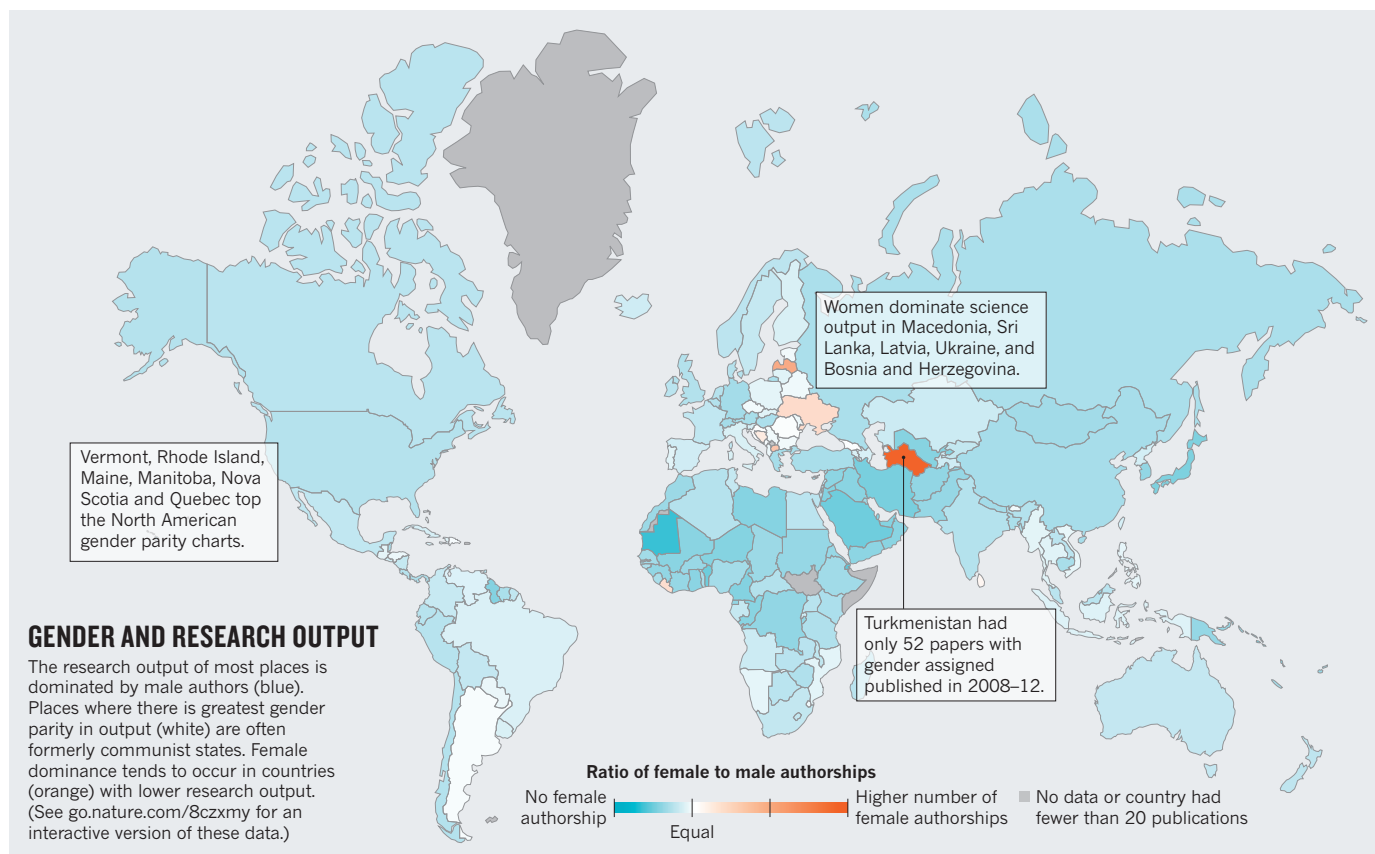
in areas in which research is expensive⁸, such as high-energy physics, possibly as a result of policies and procedures relating to funding allocations⁴. Women are less likely to participate in collaborations that lead to publication and are much less likely to be listed as either first or last author on a paper⁷. There is no consensus on the reasons for these gender differences in research output and collaboration — whether it is down to bias,

childbearing and rearing⁹, or other variables.

It has been suggested that what women lack in research output they make up for in citations, particularly in fields with ‘greater career risk’⁸ — that is, fields with long lags between doctoral education and securing a faculty position, such as ecology. But again, there is no consensus on the relative impact of women's work compared to men's.

The present state of quantitative knowledge of gender disparities in science has been shaped primarily by anecdotal reports and studies that are highly localized, ►

► **NATURE.COM**
See *Nature's* special issue on women in science:
nature.com/women



► monodisciplinary and dated. Furthermore, these studies take little account of the rise in collaborative research and other changes in scholarly practices. Effective policy cannot be built on such foundations.

Therefore, we present here a global and cross-disciplinary bibliometric analysis of: first, the relationship between gender and research output (for which our proxy was authorship on published papers); second, the extent of collaboration (for which our proxy was co-authorships); and third, scientific impact of all articles published between 2008 and 2012 and indexed in the Thomson Reuters Web of Science databases (for which our proxy was citations). We analysed 5,483,841 research papers and review articles with 27,329,915 authorships. We assigned gender using data from the US Social Security database, among other sources (see Supplementary Information; go.nature.com/j3otjz).

We find that in the most productive countries, all articles with women in dominant author positions receive fewer citations than those with men in the same positions. And this citation disadvantage is accentuated by the fact that women's publication portfolios are more domestic than their male colleagues — they profit less from the extra citations that international collaborations accrue. Given that citations now play a central part in the evaluation of researchers, this situation can only worsen gender disparities.

In our view, the scale of this study provides

much-needed empirical evidence of the inequality that is still all too pervasive in science. It should serve as a call to action for the development of higher education and science policy.

BIAS BY NUMBERS

Men dominate scientific production in nearly every country; to what extent varies by region (see 'Gender and research output'). We probed the proportion of each gender's output by comparing the proportion of identified authorships for each gender on any given paper. For example, for a paper with eight authorships, of which six were assigned a gender, each of the authorships would be granted one-sixth of a paper. These gendered fractions were then aggregated at the levels of countries and disciplines. It should be stressed that these are authorships, not individuals, therefore no author name disambiguation was necessary (see Supplementary Information).

Globally, women account for fewer than 30% of fractionalized authorships, whereas men represent slightly more than 70%. Women are similarly underrepresented when it comes to first authorships. For every article with a female first author, there are nearly two (1.93) articles first-authored by men.

South American and Eastern European countries demonstrate greater gender parity. Eastern Europe may support the idea that communist and formerly communist states

may have greater gender balance than other countries. Only nine countries had female dominance in terms of proportion of authorships, and five of these (Macedonia, Sri Lanka, Latvia, Ukraine, and Bosnia and Herzegovina) had more than 1,000 articles in our analysis. In other words, female authorship is more prevalent in countries with lower scientific output.

Countries with more than 1,000 papers and high degrees of male dominance include, unsurprisingly (in order of output): Saudi Arabia, Iran, Japan, Jordan, the United Arab Emirates, Cameroon, Qatar and Uzbekistan. US states with more than 1,000 articles with a gender assigned and high male dominance include New Mexico, Mississippi and Wyoming. The US states and Canadian provinces that are closest to achieving gender parity (and have more than 1,000 articles) include Vermont, Rhode Island, Maine, Manitoba, Nova Scotia and Quebec. Again, some of these states and provinces are among the lowest ranking in terms of scientific output.

Our disciplinary results confirmed previous findings and anecdotal knowledge about fields associated with 'care'. Specialties dominated by women include nursing; midwifery; speech, language and hearing; education; social work and librarianship. Male-dominated disciplines include military sciences, engineering, robotics, aeronautics and astronautics, high-energy physics, mathematics, computer science, philosophy

and economics. Although disciplines from the social sciences show a larger proportion of female authors, the humanities are still heavily dominated by men.

Next we looked at collaboration. We analysed the proportion of papers by gender that are the result of national collaboration, compared with those that result from international collaborations. For the 50 most productive countries in our analysis (accounting for 97% of the total publications), female collaborations are more domestically oriented than are the collaborations of males from the same country.

And what of impact? We analysed prominent author positions — sole authorship, first-authorship and last-authorship. We discovered that when a woman was in any of these roles, a paper attracted fewer citations than in cases in which a man was in one of these roles (see 'Lead-author gender and citation'). The gender disparity holds for national and international collaborations.

AGE-OLD STORY

There are several limitations to the conclusions that can be drawn from our findings. Foremost among them is that age indisputably has a role — perhaps even the major role — in explaining gender differences in scientific output, collaboration and impact. As is well known, the academic pipeline from junior to senior faculty leaks female scientists, and the senior ranks of science bear the imprint of previous generations' barriers to the progression of women. Thus it is likely that many of the trends we observed can be explained by the under-representation of women among the elders of science. After all, seniority, authorship position, collaboration and citation are all highly interlinked variables.

Another key limitation is that authorship

of papers is only one of many indicators of research activity. Our analysis includes only journal articles, not books, conference proceedings, database construction or code, for example. Also problematic is the lack of universal norms associated with authorship attribution and position. For example, it is possible that some women do not appear as authors despite their contribution to research activities, and there are fields in which authors are listed alphabetically. There is also a concern that gender-assignment techniques can introduce errors (see Supplementary Information). We have tried to mitigate this with validation exercises, but there is always room for improvement.

Future research should drill into questions raised by this analysis. What distinguishes pockets of anomalously high parity? Are there characteristics of the work itself that contribute to disparities in output and citation? Are there other, perhaps less quantitative, aspects of scholarship that reveal a different story regarding gender balance in science? Furthermore, is there anything intrinsic to certain disciplines or cultures that make them more or less appealing to scientists of a particular gender?

LEVELLING THE PLAYING FIELD

Those of a misogynistic bent might read this study as confirming their view that women's research is weaker than men's and there is less of it. Such a simplistic interpretation dismisses the vast implications embedded in these data. Our study lends solid quantitative support to what is intuitively known: barriers

"Programmes fostering international collaboration for female researchers might help to level the playing field."

to women in science remain widespread worldwide, despite more than a decade of policies aimed at levelling the playing field. UNESCO data show¹⁰ that in 17% of countries an equal number of men and women are scientists. Yet we found a grimmer picture: fewer than 6% of countries represented in the Web of Science come close to achieving gender parity in terms of papers published.

For a country to be scientifically competitive, it needs to maximize its human intellectual capital. Our data suggest that, because collaboration is one of the main drivers of research output and scientific impact, programmes fostering international collaboration for female researchers might help to level the playing field.

That said, if there were a simple solution or programme that could improve matters, this issue would already be solved. Unfortunately, behind this global imbalance lie local and historical forces that subtly contribute to the systemic inequalities that hinder women's access to and progress in science. Any realistic policy to enhance women's participation in the scientific workforce must take into account the variety of social, cultural, economic and political contexts in which students learn science and scientific work is performed. Each country should carefully identify the micro-mechanisms that contribute to reproducing the past order. No country can afford to neglect the intellectual contributions of half its population. ■

Vincent Larivière is chaired assistant professor of transformations of scholarly communication at the University of Montreal, Canada. **Chaoqun Ni** is in the School of Informatics and Computing at Indiana University Bloomington.

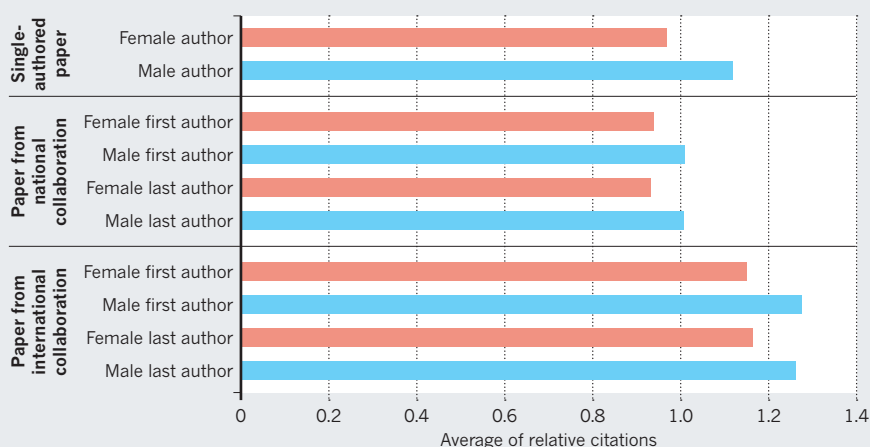
Yves Gingras is chaired professor of history and sociology of science at the University of Quebec at Montreal, Canada.

Blaise Cronin is chaired professor of information science at Indiana University Bloomington. **Cassidy R. Sugimoto** is assistant professor in the School of Informatics and Computing, Indiana University Bloomington. e-mail: sugimoto@indiana.edu.

1. Organisation for Economic Co-operation and Development. *Education at a Glance 2012* (OECD, 2012).
2. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. *Proc. Natl Acad. Sci. USA* **109**, 16474–16479 (2012).
3. Shen, H. *Nature* **495**, 22–24 (2013).
4. Ley, T. J. & Hamilton B. H. *Science* **322**, 1472–1474 (2008).
5. Holden, C. *Science* **294**, 396–411 (2001).
6. Ding, W. W. *Science* **313**, 665–667 (2006).
7. West, J. D., Jacquet, J., King, M., Correll, S. J. & Bergstrom, C. T. *PLoS ONE* e66212 (2013).
8. Duch, J. et al. *PLoS ONE* **7**, e51332 (2012).
9. Ceci, S. J. & Williams, W. M. *Proc. Natl Acad. Sci. USA* **108**, 3157–3162 (2011).
10. United Nations Educational, Scientific and Cultural Organization. *Science, Technology and Gender: An International Report* (UNESCO, 2007).

LEAD-AUTHOR GENDER AND CITATION

Papers with female authors in key positions are cited less than those with male authors in key positions, be they papers with one author, or those resulting from national or international collaborations.



➔ FOR AN INTERACTIVE VERSION OF THIS CHART SEE: go.nature.com/j3otjz



The computer-generated Gollum (left, seen in 2012's *The Hobbit: An Unexpected Journey*) is based on performances by actor Andy Serkis.

COMPUTER ANIMATION

Digital heroes and computer-generated worlds

Joe Letteri, Oscar-winning co-creator of Gollum on screen, looks at the evolution of computer animation as the next instalment of *The Hobbit* trilogy reaches cinemas.

From the beginning, computer animation has had the ability to engage viewers by giving artists a way to mix reality with fantasy. Early film examples such as the stained-glass knight in Barry Levinson's 1985 *Young Sherlock Holmes* and the 'water tentacle' in James Cameron's 1989 *The Abyss* offered a glimpse of the potential of this new art form to create memorable characters. That became immediately apparent in Cameron's later *Terminator 2* (1991). He took an already great idea for a character (a terminator robot) and turned it in a new and unexpected direction: the shape-shifting T-1000 liquid-metal terminator.

Computer animation is a natural extension of hand-drawn methods developed during the early-twentieth-century golden age of 'cel' animation ushered in by Walt Disney, in which a series of images is played back at speed to give the illusion of life. Just

as we do today, animators used a variety of reference techniques to capture the essence of organic movement. Snow White's dance with the dwarves in Disney's 1937 film was created by matching the movements of a live, filmed dancer, using a technique called rotoscoping — in the most basic terms, tracing the motion from a film one frame at a time. This technique, although now slightly more sophisticated in its application, is still in use.

Take dinosaurs, for example. I was fortunate to begin my career at the US visual-effects company Industrial Light & Magic, then in Marin County, California, as it was gearing up to create the computer-generated dinosaurs for Steven Spielberg's 1993 *Jurassic Park*. Naturally, we all thought that rotoscoping dinosaurs would be a great idea, but unfortunately that

was out of the question. Instead, we studied elephants to understand weight, and lizards, other reptiles and birds to get some ideas about how dinosaurs of different sizes might have moved. Digital animators did motion studies, copying the movement of these animals frame by frame until they could synthesize a convincing idea of dinosaur movement.

Two years later, computer animation took another big step forward with the astounding success of Pixar's *Toy Story*. Software was becoming sophisticated enough to tackle the creation of a character's performance. In traditional animation, a lead animator sets key frames or poses for a character and junior animators draw the 'in-betweens' from pose to pose. Now, artists could use the computer to do that. Pixar proved that three-dimensional computer animation could be used to create an entire film.

NATURE.COM
For a video on the making of Gollum:
go.nature.com/7qtgxr



In 2001, I joined Wellington-based visual-effects company Weta Digital to work with Peter Jackson on *The Lord of the Rings* trilogy, inspired in large part by the chance to create the character Gollum. Gollum was a special challenge, because the more realistically human a character is, the more complex the animation gets. People are attuned to recognizing all aspects of human motion and behaviour, no matter how subtle. And because the characters we create are three-dimensional, we have to understand how to pose them frame by frame to achieve realistic performances.

With Gollum, we used the then relatively new technique of performance capture — effectively, an extension of rotoscoping. But instead of looking at an actor from a single point of view and matching the form of the motion, we looked at his performance using dozens of cameras simultaneously and

matched the motion's underlying dynamics.

Andy Serkis, who performed Gollum, wore a special suit with reflective markers to show the key positions of his joints. From the multiple cameras, we could calculate his skeleton's position at every frame as he performed. Those positions were then transferred to Gollum's digital skeleton, which allowed us to make the character move the way Andy did. Traditional key-frame animation techniques still apply, however. For example, the first time we see Gollum in *The Lord of the Rings: The Two Towers* (2002), he is climbing down a vertical rock face, something no human can do. So that motion is based on animators observing what a human can do and using their imaginations to create a believable performance. And, in a direct throwback to rotoscoping, we created his facial performance and dialogue-related movements by hand, frame by frame, from Andy's filmed performance.

The problem with trying to capture facial motion is that there are no joints, apart from the jaw, that have movements you can track. So for Jackson's 2005 *King Kong*, Weta came up with a different technique. Again with Serkis, we glued small reflective markers all over his face. By using these to track the changes in skin position and tension as Andy performed, we could compute what his muscles were doing

underneath. Then we built Kong so that he had the same facial-muscle layout as Andy, and used Andy's muscle movements to drive Kong's facial performance.

This breakthrough meant that we could now capture an actor's performance in its entirety. This became important for Weta's next film, James Cameron's *Avatar* (2009), for which we made one important modification. Each actor wore a helmet that filmed their facial movements; we then extracted the performance data from each frame and used a 'facial action coding system' solver to translate the movements into muscle activations. The knowledge of which muscles are activated in a facial expression informed our activation of the corresponding muscles in the digital characters. In addition, this process allowed the director to see the actors live through a virtual camera as they were instantly transformed into their Na'vi characters moving through the world of Pandora.

Digital characters also have to appear realistic in their surroundings, whether that is a photographed environment or a complete digital creation such as the jungles of Pandora. So we looked to understand how light and materials interact in nature. One of the best examples of this interaction is subsurface scattering. We first developed a technique to replicate this mechanism of light transport to create the translucency of Gollum's skin, leveraging pioneering research by computer-graphics specialist Henrik Wann Jensen and his colleagues at Stanford University in California (see go.nature.com/lyzuh2). The ▶



The Na'vi in *Avatar* (2009) lived in an entirely digitally created world.

► thick skin of a dinosaur can be simulated by bouncing light off the exterior. But human skin is softer and more translucent, so light enters and bounces around dozens of times before exiting. These properties, which are easily observed by putting your hand in front of a bright light, are crucial to a realistic portrayal.

Realistic animation also depends on knowledge of how skin, muscles and hair move independently of a character's performance.

"Realistic animation depends on knowledge of how skin, muscles and hair move independently of a character's performance."

These secondary motions are achieved through intensive simulations that compute all of the mass, dynamics, tensions and interaction of each part of the body as a character moves.

The simulations help to create the complex visual cues that the human brain processes when taking in an image. They also ensure that the physiology of creatures (real or fantastical) has a ground truth and is believable. Combining this new level of detail with motion-captured performances of talented actors has enabled computer-animated characters to become lead actors.

Rupert Wyatt's 2011 *Rise of the Planet of the Apes* featured a chimpanzee named Caesar who was not just the protagonist; he was the emotional centre of the film. This entirely digital character is a great example of how the advances in animation work together, from muscle simulation, fur and realistic lighting, to motion-captured body and facial performance.

The following year saw all of these developments come full circle when we were able to once again present Gollum in Peter Jackson's *The Hobbit: An Unexpected Journey*. This new Gollum benefited from a much more detailed digital model, new subsurface scattering techniques and all of the advances we have made in the past ten years. If you watch closely, you will see the muscles moving under his skin and the light refracting in his eyes. And you will get a glimpse of the worlds we can create from the mix of all this art and science. ■

Joe Letteri is senior visual effects supervisor at Weta Digital in Wellington, New Zealand. He has received four Academy Awards for Visual Effects and the Academy's Technical Achievement Award for co-developing the subsurface-scattering technique that brought Gollum to life. His latest film is *The Hobbit: The Desolation of Smaug*, on general release this week. e-mail: digital@wetafx.co.nz



John Milne (left) with his wife Tone, seismologist Boris Galitzin and a lamp-post seismometer in 1910.

GEOLOGY

The maverick founder of modern seismology

George Helffrich relishes a film on John Milne, whose work in Japan put earthquake science on the map.

What European other than Marco Polo, on taking up a job in the Far East, would travel overland? John Milne. No posh Brit he: the Victorian geologist reached Japan mainly by train, foot and pack animal, ostensibly to avoid seasickness. As I write this, flying to Japan across eastern Siberia, I am amazed. I detect a whiff of the iconoclast.

Milne's motivation for that epic journey was an invitation from Japanese officials of the Meiji era (1868–1912) to establish an Imperial College of Engineering in Tokyo, and to transfer Western knowledge to Japan. After arriving in 1875 to teach geology and mining, he became aware of the frequency of earthquakes and their damaging effects. Building designs transplanted from the West performed particularly poorly. By 1878, Milne determined that to study earthquake damage properly, he needed to quantify seismicity — or the frequency and strength

of earthquakes — instrumentally. His goal became to build a standard seismometer and establish a seismic network across Japan. That plan was the inception of modern global seismology.

To mark this year's centenary of Milne's death, his great-nephew William Twycross made a documentary, *The Man Who Mapped the Shaking Earth*. Shown in July at the assembly of the International Union of Geodesy and Geophysics in Gothenburg, Sweden, the film immerses us in Milne's life in science, including his photographs and sketches. Often shooting on location, Twycross and his film crew trace Milne's travels in Iceland, the Canadian island of Newfoundland, Britain, Russia and the United States, as well as to and in Japan. Twycross imposes a fine narrative continuity

The Man Who Mapped the Shaking Earth
DIRECTOR: WILLIAM TWYXCROSS
2013.

despite the logistical complexities.

Milne's astonishing saga began at King's College London, where he studied science. His published account of a study trip to Iceland in 1871, when he was 21, showed him to be a competent diarist and illustrator, and won him a scholarship to the Royal School of Mines (now part of Imperial College London). He suspended his studies there to complete a commercial geological survey of Newfoundland, where he incidentally studied the recently extinct great auk. On a Royal Geographical Society expedition to Egypt's Sinai peninsula in 1874, he marshalled his new skills in geology and illustration to create geographical cross-sections of the region — work that paved the way for his post in Japan.

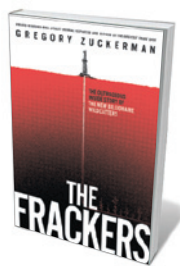
Once there, he published his trans-Asian travelogue in the *Transactions of the Asiatic Society of Japan*, and climbed, studied, sketched and even discovered some volcanoes, one of which is now named Milne. He made archaeological and anthropological studies of the indigenous Ainu culture and people of the northern Japanese islands, and met his wife-to-be, Tone Horikawa. After the Tokyo–Yokohama earthquake of 1880, he established the Seismological Society of Japan and its journal — both world firsts. The Royal Society of London elected him a fellow in 1887.

In 1895, Milne left Japan with Tone for the Isle of Wight, off the south coast of Britain, building a seismological observatory there and working towards the standardization of instruments such as the horizontal pendulum seismograph to record earthquakes worldwide. The captaincy of the island's Newport Golf Club and membership of the local photography and chess clubs occupied his idle moments. By 1903 he had set up the world's first global seismic network, with 40 stations encompassing every continent reporting to his observatory, Shide Hill House. He also published the periodic *Shide Circular Reports on Earthquakes* from 1900 to 1912, a precursor to the seismological catalogues accumulated by national agencies around the world. His obituary (see *Nature* **91**, 587–588; 1913) hints at some resistance to coordinating his network with the International Seismological Association — not surprising, given his personal achievement.

The Man Who Mapped the Shaking Earth paints an unembellished portrait of a dedicated visionary. Milne wove the threads of seismic instrumentation and observation into a tool to develop fundamental knowledge of how Earth works. What could be more compelling or inspiring to a student of seismology or the history of science? ■

George Helffrich is professor of seismology at the University of Bristol School of Earth Sciences, UK.
e-mail: george.helffrich@bristol.ac.uk

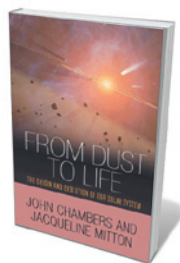
Books in brief



The Frackers: The Outrageous Inside Story of the New Billionaire Wildcatters

Gregory Zuckerman PORTFOLIO (2013)

Investigative journalist Gregory Zuckerman maps the landscape of the new US oil boom, currently producing 7.5 million barrels of crude a day. Focusing on the widespread technique of fracking — hydraulic fracturing of deep shale deposits for oil and gas — Zuckerman plaits portraits of the 'wildcatters' who champion the practice into a pacy chronicle of the boom. His take on the issues is balanced, but this is not the place to find a lengthy analysis of the potential impacts of fracking on environmental systems.



From Dust to Life: The Origin and Evolution of Our Solar System

John Chambers and Jacqueline Mitton PRINCETON UNIVERSITY PRESS (2013)

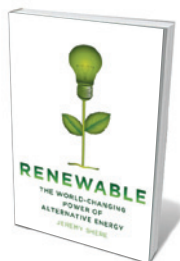
This research round-up of how our "cosmic neighbourhood" evolved is a stellar read. Planetary scientist John Chambers and science writer Jacqueline Mitton shuttle the reader through the big discoveries — from heliocentricity to findings in the past 50 years, as missions, telescopes and light detectors probe deeper. Lingering mysteries such as early planet formation are laid out, along with details on current missions such as New Horizons, headed for the Kuiper belt of asteroids and set to encounter Pluto in July 2015.



Ship of Death: A Voyage That Changed the Atlantic World

Billy G. Smith YALE UNIVERSITY PRESS (2013)

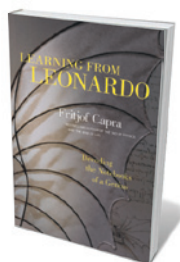
One of the strangest tales in the annals of medicine must be the voyage of the *Hankey*. Historian Billy G. Smith has resurrected the long-buried history of the British abolitionists who set off in the ship to found a slavery-free West African colony in the late eighteenth century. They failed — then inadvertently wreaked viral havoc as the *Hankey* left Africa riddled with mosquitoes carrying yellow fever. Stopping at various Atlantic ports, the ship contributed to a pandemic that killed tens of thousands. Packed with fascinating nuggets — such as the part played by that deadly cargo in the liberation of Haiti.



Renewable: The World-Changing Power of Alternative Energy

Jeremy Shere ST MARTIN'S PRESS (2013)

Renewable energy is a tortuous, fast-moving field, and in this contribution to the shelves of books on it, science journalist Jeremy Shere offers a welcome pragmatism. Organizing his survey by power source, from plants to water, he examines state-of-the-art technologies and their surprising histories. We visit researchers who probe cellulosic energy crops, algal oil, nanotechnology-enhanced solar cells and much more. At a time when countries including the United States and China are investing heavily in renewables, Shere notes that combining alternative-energy choices strategically is key.



Learning from Leonardo: Decoding the Notebooks of a Genius

Fritjof Capra BERRETT-KOEHLER PUBLISHERS (2013)

Was Leonardo da Vinci essentially a systems theorist? So argues physicist Fritjof Capra. His compelling exploration of Leonardo's breakthroughs in areas such as fluid dynamics reveals an intellectually fearless mind constantly seeking patterns, relationships and context. Leonardo could have changed the course of science through discoveries such as how the heart functions; but he never published, and the credit went to others. **Barbara Kiser**

Faster than kiss-and-run

An ultrafast mode of vesicle endocytosis — a crucial process occurring at neural junctions that underpins brain function — has been uncovered. Long-standing models of endocytosis will therefore need to be re-evaluated. [SEE ARTICLE P.242](#)

SOYOUN CHO &
HENRIQUE VON GERSDORFF

The rapid propagation and exchange of signals between the neurons in our brains conveys information about sensory stimuli and underpins our thoughts, dreams and memories. This signalling requires the constant formation, absorption and recycling of vesicles that secrete neurotransmitter molecules at the junctions between neurons (synapses). On page 242 of this issue, Watanabe *et al.*¹ report that an ingenious combination of techniques has allowed them to identify a previously unknown form of endocytosis — the fundamentally important process by which the vesicular membranes are recycled. The results are exciting, compelling and totally unexpected. Furthermore, the authors' technical innovations promise to reveal the inner workings of vesicle recycling at unprecedented temporal resolution.

Synapses are the primary sites for the rapid, point-to-point communication that occurs between interconnected neurons. Most synapses in the mammalian brain consist of small, presynaptic nerve terminals (less than one micrometre in diameter) that contact the membranes of postsynaptic neurons. These 'bouton-type' terminals contain approximately 200 synaptic vesicles (30–40 nanometres across), each packed with about 2,000 neurotransmitter molecules.

About five to ten vesicles dock at a specialized membrane area called the active zone (Fig. 1). It is thought that one to three of those docked vesicles are primed to release neurotransmitters by a process called exocytosis. The arrival of a nerve impulse (an action potential) at the presynaptic terminal depolarizes the cell membrane. This causes voltage-gated calcium channels to open, and

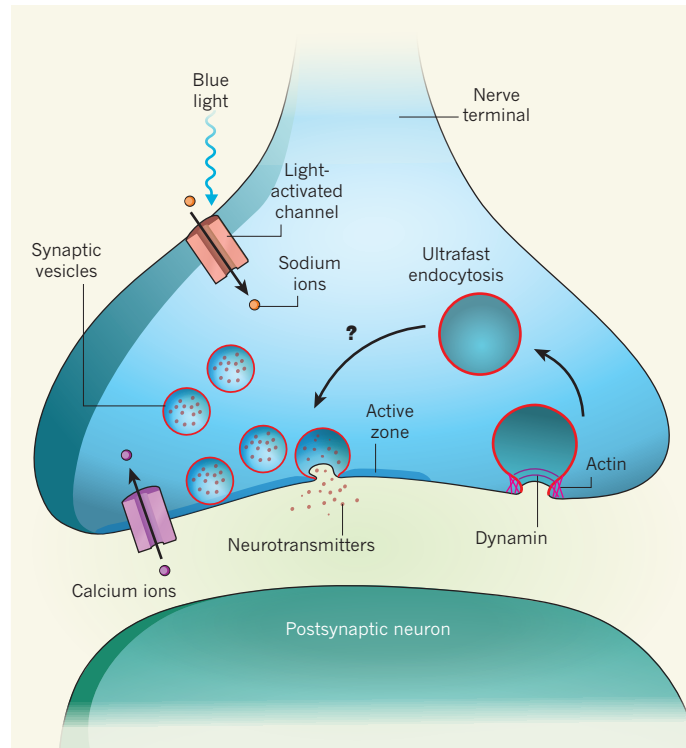


Figure 1 | Ultrafast endocytosis. Watanabe *et al.*¹ expressed light-activated channels in the cell membranes of nerve terminals *in vitro*. The channels respond to a 10-millisecond flash of blue light by allowing an influx of sodium ions. This depolarizes the cell membrane, triggering action potentials at the end of the light flash that open calcium channels in the active zone of the nerve terminal. The resulting influx of calcium ions causes docked synaptic vesicles to fuse with the cell membrane at the active zone of the terminal, releasing neurotransmitters onto the membrane of the postsynaptic neuron. This chain of events occurs within 1–2 ms. The authors observed that an ultrafast mode of endocytosis occurs within 50–100 ms after the onset of the light flash, in which large vesicles are generated from the cell membrane. The process required actin and dynamin proteins. It is not known how small synaptic vesicles are re-formed from these larger vesicles.

the subsequent influx of calcium ions triggers the fusion of docked synaptic vesicles with the membrane. The vesicles then release their cargo of neurotransmitters onto the membrane of the postsynaptic neuron. In this way, signals are passed between neurons.

Endocytosis is required to recycle fused vesicular membranes, restore the terminal's original surface area and prevent depletion of vesicle pools. Fast and slow modes of endocytosis coexist at several different types of nerve terminal, and the kinetics of the

process seems to depend on how much exocytosis has previously occurred. The slow mode tends to dominate after mild to strong stimulation and takes, on average, 10–20 seconds. This mode is thought to occur after complete collapse of the vesicle membrane into the terminal's membrane and to require several proteins, such as clathrin and dynamin, to recycle the membrane at sites outside the active zone².

By contrast, the fast mode of endocytosis prevails after weak to mild stimulation and takes, on average, only 0.3–1 seconds. The mechanisms that mediate this mode are not well understood, but it has been observed in both small and large nerve terminals^{3–6}, and seems to be most prominent and robust in mature nerve terminals^{7,8}. One proposed mechanism suggests that vesicles release neurotransmitters through a transiently opened fusion pore; the vesicle then re-forms in less than 1 second as the pore closes, and detaches from the membrane^{3,4}. Therefore, the vesicle membrane does not collapse into the terminal's membrane. This mechanism has rather whimsically been called kiss-and-run.

A third mode of endocytosis predominates during prolonged, strong stimulation that leads to the fusion of several vesicles and

occurs about 1–2 seconds after the onset of stimulation. In this bulk mode of endocytosis, a large membrane invagination is incorporated into the terminal from the cell membrane, from which several clathrin-coated vesicles can bud off².

In contrast to endocytosis, exocytosis is extremely fast: docked synaptic vesicles are thought to open their fusion pores and fully collapse within 5–20 milliseconds of calcium-channel opening. So it has been extremely hard to capture images of omega figures — synaptic

vesicles with open fusion pores before full collapse — using conventional electron microscopy (EM). However, Watanabe and colleagues' experimental approach allowed them to capture hundreds of images of EM omega figures.

The authors cultured mouse hippocampal neurons and then expressed a light-activated channel protein in their membranes by infecting them with non-toxic viruses that contained the channel's DNA. When irradiated with a 10-millisecond flash of blue light, these proteins allow sodium ions to enter the neurons (Fig. 1). This influx depolarizes nerve terminals and, at the end of the light flash, triggers one or two action potentials, which open voltage-gated calcium channels at active zones. The resulting influx of calcium ions then triggers the fusion of docked synaptic vesicles with the plasma membrane and the release of neurotransmitters.

Because this chain of events occurs within a few milliseconds, Watanabe and co-workers engineered a device that cooled the nerve terminals from about 34 °C to roughly 0 °C as early as 10 ms after the initiation of vesicle fusion. This rapid freezing process allowed the researchers to capture EM images that directly confirm the long-standing hypothesis that a subset of primed and docked vesicles fuse preferentially at active zones rather than outside them. Furthermore, the authors observed that vesicle-depleted active zones are repopulated with a full complement of docked vesicles within roughly 4–10 seconds. This coincides with the average time course for recovery from synaptic depression (the activity-dependent, temporary inability of a neuron to secrete neurotransmitters), which suggests that the depletion of a primed subset of docked vesicles is one major mechanism for short-term synaptic depression.

Watanabe and colleagues went on to capture morphological changes in the nerve terminal's cell membrane at several time points after light-evoked stimulation of the terminal. They observed that, after exocytosis, the vesicular membrane fully collapses and flattens onto the plasma membrane. About 50 ms after the light flash, the first signs of membrane invagination (the beginnings of endocytosis) occur at locations outside the active zone. As early as 50–100 ms after the onset of the light flash, the authors observed large invaginations that seemed to form unusually large vesicles (Fig. 1). These vesicles have a four-fold greater surface area than normal synaptic vesicles and do not seem to have standard clathrin coats. Therefore, this ultrafast mode of endocytosis is mediated by an unidentified set of molecules, but is clearly distinct from kiss-and-run because it occurs outside the active zone.

What is the signal that leads to the rapid formation of these large vesicles? One possibility is the sudden reduction in membrane

tension that occurs after the full fusion and collapse of docked vesicles. This reduction could lead to the rapid invagination of previously fused synaptic-vesicle membranes. Indeed, fast endocytosis is blocked by an increase in neuronal hydrostatic pressure and membrane tension at some nerve terminals⁹.

Watanabe and colleagues' results constitute a major technical advance over earlier work because they reveal clear EM omega figures after only one or two action potentials. EM omega figures could previously be captured routinely only under non-physiological conditions in which synapses were treated with drugs to prolong action-potential duration and then stimulated. Nevertheless, earlier pioneering experiments¹⁰ using the rapid-freezing technique were the first to prove that neurotransmitter release occurs through the exocytosis of synaptic vesicles. The combination of optical stimulation with rapid-freezing EM now promises to reveal the details of the first few hundred milliseconds after exocytosis at small bouton-type nerve terminals in the mammalian central nervous system. As we zoom in on the intricate nanomachinery that

couples exocytosis to endocytosis with ever-higher temporal resolution, we are reminded that a pristine picture is often worth more than a thousand words. ■

Soyoun Cho and Henrique von Gersdorff
are at the Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239, USA.
e-mail: vongersd@ohsu.edu

1. Watanabe, S. *et al. Nature* **504**, 242–247 (2013).
2. Saheki, Y. & De Camilli, P. *Cold Spring Harb. Perspect. Biol.* **4**, a005645 (2012).
3. Alabi, A. A. & Tsien, R. W. *Annu. Rev. Physiol.* **75**, 393–422 (2013).
4. Gandhi, S. P. & Stevens, C. F. *Nature* **423**, 607–613 (2003).
5. von Gersdorff, H. & Matthews, G. *Nature* **367**, 735–739 (1994).
6. He, L., Wu, X.-S., Mohan, R. & Wu, L.-G. *Nature* **444**, 102–105 (2006).
7. Renden, R. & von Gersdorff, H. *J. Neurophysiol.* **98**, 3349–3359 (2007).
8. Rose, T., Schoenenberger, P., Jezek, K. & Oertner, T. G. *Neuron* **77**, 1109–1121 (2013).
9. Heidelberger, R., Zhou, Z.-Y. & Matthews, G. *J. Neurophysiol.* **88**, 2509–2517 (2002).
10. Heuser, J. E. *Q. J. Exp. Physiol.* **74**, 1051–1069 (1989).

This article was published online on 4 December 2013.

EXTRASOLAR PLANETS

Inner edge of the habitable zone

A three-dimensional climate model indicates that the fraction of Sun-like stars that might harbour a rocky planet within their habitable zone could be smaller than previously estimated. [SEE LETTER P.268](#)

JAMES F. KASTING & CHESTER E. HARMAN

The successful, if perhaps somewhat premature, conclusion of NASA's Kepler mission has spawned renewed interest in potentially habitable planets. According to recent research¹, some 22% of Sun-like stars might host a rocky planet within their habitable zone — the region within which liquid water can exist on the planet's surface. With roughly 100 billion stars in the Galaxy, this implies that there could be as many as 22 billion other Earth-like planets. But this estimate depends crucially on how the zone's boundaries are defined. The habitable zone has recently been redefined¹ as extending from 0.5 to 2 astronomical units around a Sun-like star, where 1 AU is the mean Earth–Sun distance. This is a curious choice, because the completely desiccated planet Venus orbits at 0.72 AU, suggesting that the habitable-zone inner edge is somewhere beyond that distance. On page 268 of this issue, Leconte *et al.*² show that the inner edge probably lies closer to

0.95 AU, where the solar flux is about 1.1 times that at Earth's orbit (Box 1).

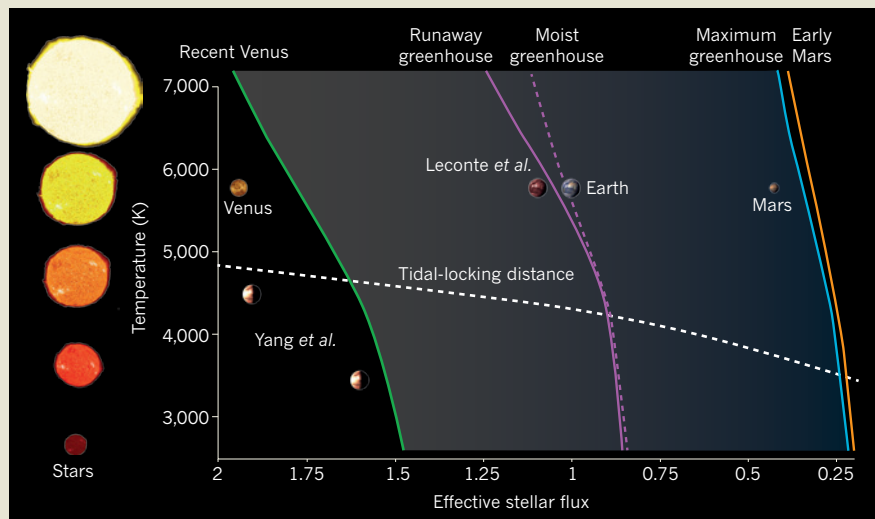
If Leconte and colleagues' result is correct, and if planets are spaced as they are in our Solar System, then the previous estimate¹ for the fraction of Sun-like stars that might have a rocky planet within their habitable zone — which astronomers term η_{Earth} ('eta-sub-Earth') — could be too high by a factor of almost 2. This difference is important because a smaller value for η_{Earth} implies that a future space telescope designed to observe and characterize such planets must be correspondingly larger.

Ironically, the new estimate of 0.95 AU is essentially the same as two older estimates^{3,4} for the inner edge of the habitable zone. Twenty years ago, researchers calculated³ that this orbital distance was where a moist-greenhouse atmosphere would occur. This is an atmosphere in which liquid water remains present at the planet's surface, but the stratosphere, which lies just above the lowest portion of the atmosphere, is also wet. This allows

BOX 1

Habitable zones for stars of various surface temperatures

The horizontal axis represents the stellar flux at a planet's orbit relative to the flux at Earth's orbit (effective stellar flux). The coloured curves show different estimates for the inner and outer edges of the habitable zone⁶. Three estimates are given for the inner edge: the 'moist-greenhouse' limit is where a planet's stratosphere becomes wet, so that water is lost by photodissociation and hydrogen escape; the 'runaway-greenhouse' limit is where all surface water evaporates; and the 'recent-Venus' limit is based on the observation that Venus seems to have lost its water by 1 billion years (Gyr) ago, when the Sun was 8% less bright than today^{3,6}. Two estimates are shown for the outer edge: the 'maximum-greenhouse' limit is the theoretical limit for a CO₂-H₂O atmosphere, based on 1D model calculations; the 'early-Mars' limit is based on the observation that Mars could have been habitable 3.8 Gyr ago, when the Sun was 25% less bright^{3,6}. These curves move towards lower stellar flux at lower stellar surface temperatures, because such stars emit much of their radiation at long wavelengths, causing their planets' albedos to be lower. The white dashed curve corresponds to the orbital distance within



which an Earth-like planet with an initial rotational period of 13.5 hours would become tidally locked within 4.5 Gyr (ref. 9). Planets older than 4.5 Gyr and lying below this curve might end up in synchronous rotation, as the Moon is with Earth. The reddish planet just inside Earth's orbit shows the position of the habitable-zone inner edge as calculated by Leconte *et al.*² using their 3D climate

model. The two half-illuminated planets towards the lower left show the position of the inner edge for two synchronously rotating planets simulated by Yang *et al.*⁸ earlier this year using a different 3D model. Planets orbiting such dim red stars could conceivably remain habitable even if they absorb more stellar radiation than Venus. **J.F.K. & C.E.H.**

water vapour to be photodissociated by solar ultraviolet radiation, with the resulting hydrogen escaping efficiently to space. Long before that, it had been estimated⁴ that a runaway-greenhouse atmosphere — one in which all water is present in the atmosphere as steam — would occur at that same orbital distance. Both phenomena would make life impossible on the planet's surface, either immediately (for the runaway greenhouse) or at some time in the future (for the moist greenhouse). These two earlier studies were based on one-dimensional (1D), globally averaged climate models. By contrast, Leconte *et al.* use a 3D climate model, which enables them to account for phenomena such as clouds and relative humidity variations that are typically neglected in 1D models.

That Leconte and colleagues' inner-edge estimate agrees with the two older values is fortuitous, because several factors have changed during this time. Most importantly, the efficiency with which water vapour absorbs visible radiation (sunlight) is now thought to be higher than previously assumed^{5,6}. This lowers a water-rich planet's albedo, or reflectance, causing the distance at which a moist greenhouse occurs in 1D models to move out to 0.99 AU, perilously close to Earth's orbit⁶. But the more sophisticated³ of the two earlier

1D models assumes that the lowest portion of the atmosphere, the troposphere, is fully saturated, whereas the 3D model does not, and so the inner edge in the 3D model is pushed back towards the Sun.

Somewhat surprisingly, clouds in the 3D model are predicted to destabilize Earth's climate, because high cirrus clouds (which warm the surface) increase in number and size faster at high surface temperatures than do low stratus clouds (which cool it). The same type of positive cloud feedback is observed in many studies of doubled carbon dioxide levels on modern Earth⁷; however, previous workers (see, for example, ref. 3) have consistently assumed that cloud feedback would be negative in dense, water-rich atmospheres.

The cloud-feedback story has further twists. Earlier this year, Yang *et al.*⁸ used a different 3D climate model to calculate where the habitable-zone inner edge lies for synchronously rotating (tidally locked) planets around M- and late K-type stars, which have lower temperatures than the Sun. (Tidal locking refers to the slowing of a planet's rotation rate by tides raised by its parent star.) This study⁸ found that the sunlit sides of such planets were perpetually cloudy and that the clouds reduced the surface temperature by increasing the planet's albedo. If so, then the cloud feedback for such planets

is negative, not positive. But the atmospheric circulation patterns on slowly rotating, tidally locked planets are quite different from those on rapidly rotating planets such as Earth, and so it may be that clouds simply behave differently in this situation.

One further prediction of Leconte and colleagues' model should be noted. At warm surface temperatures, 330 kelvin or above, the model's predicted temperature at a planet's tropopause is quite low, about 120 K (the tropopause is the upper boundary of the troposphere). The tropopause in this model is located roughly 80 kilometres above the surface for the warmest atmosphere considered (see Fig. 3 of the paper). A temperature of 120 K is cold enough to condense out most of the water vapour, allowing the atmospheric 'cold trap' to remain in operation. The cold trap on modern Earth, which limits the amount of water vapour that makes it into the stratosphere, is located at an altitude of roughly 10–15 km.

By contrast, the earlier 1D model of ref. 3 assumed a fixed temperature for the tropopause of 200 K, regardless of the planet's surface temperature, making the cold trap much less effective. A temperature of 200 K is close to that near the top of a grey atmosphere (an atmosphere that absorbs infrared radiation

equally efficiently at all wavelengths), and so this assumption seemed reasonable at the time. However, Leconte *et al.* argue that this constraint does not apply when the ability of the atmosphere to absorb infrared radiation varies with wavelength, as it does in reality. Consequently, their model does not find a moist-greenhouse limit to the habitable zone, only the runaway-greenhouse limit. Further calculations may be needed to determine whether this result is robust. ■

James F. Kasting and Chester E. Harman are in the Department of Geosciences, Penn State University, University Park, Pennsylvania 16802, USA.
e-mail: jfk4@psu.edu

1. Petigura, E. A., Howard, A. W. & Marcy, G. W. *Proc. Natl Acad. Sci. USA* **110**, 19273–19278 (2013).
2. Leconte, J., Forget, F., Charnay, B., Wordsworth, R. & Pottier, A. *Nature* **504**, 268–271 (2013).
3. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. *Icarus* **101**, 108–128 (1993).
4. Hart, M. H. *Icarus* **33**, 23–39 (1978).

5. Goldblatt, C., Robinson, T. D., Zahnle, K. J. & Crisp, D. *Nature Geosci.* **6**, 661–667 (2013).
6. Koppapur, R. K. *et al. Astrophys. J.* **765**, 131 (2013).
7. Webb, M. *et al. The Cloud Feedback Model Intercomparison Project (CFMIP): Progress and Future Plans* (Met Office, 2006); <http://ceres.larc.nasa.gov/documents/STM/2006-10/0610261450Webb.pdf>
8. Yang, J., Cowan, N. B. & Abbot, D. S. *Astrophys. J.* **771**, L45 (2013).
9. Edson, A., Lee, S., Bannon, P., Kasting, J. F. & Pollard, D. *Icarus* **212**, 1–13 (2011).

DEVELOPMENTAL BIOLOGY

Mechanics in the embryo

Mechanical stresses that are intrinsic to the early shape-forming movements of embryonic tissues have now been shown to play essential and evolutionarily conserved parts in cell-fate specification.

STEFANO PICCOLO

There is a time in animal development when the embryo, until then a single layer of identical cells, starts to fold on itself. This remarkable process, called gastrulation, produces three germ layers: the mesoderm, ectoderm and endoderm¹. This embryonic structure was already present in Urbilateria², the most recent common ancestor of most animals, which lived in the Precambrian seas more than 600 million years ago. Surprisingly, however, the cellular and molecular events associated with gastrulation are not evolutionarily conserved among extant animal phyla³. How can this universal morphogenetic step occur without any apparent selective pressure for specific mechanisms? Writing in *Nature Communications*, Brunet *et al.*⁴ suggest a solution to this conundrum — that it is the mechanics of gastrulation that is evolutionarily conserved.

During gastrulation, the mesoderm (from which muscles, bones, blood vessels and connective tissue derive) forms between the internal gut tube (the endoderm) and the ectoderm, which faces the external world and becomes neural tissue and skin. Gastrulation is a profoundly physical event: cells deform, get pulled and are compressed, generating hinges, folds and borders. But it comes in many variations. In arthropods, for example, the single-layer embryo invaginates to make a pocket that becomes the mesoderm, whereas in fish and amphibians gastrulation entails a mass movement of cells called epiboly (Fig. 1a). The molecular signals that trigger these events also differ among animal phyla.

Developmental biologists are well equipped

to play with genes and biochemical pathways, but much less at ease studying the mechanical stresses to which cells are exposed. Brunet and colleagues used a suite of tools and new bioassays to manipulate the mechanics of zebrafish embryos *in vivo*. Cell mechanics is

rooted in cytoskeletal tension and architecture, and tuned by the physical resistance of surrounding cells or the extracellular matrix⁵. By applying cytoskeleton-targeting drugs that cause epiboly to be arrested at very early stages (Fig. 1b), they found that this mechanical impairment had no effect on initial germ-layer specification and dorso-ventral (front-to-back) polarity, but that expression of the *ntl* gene, an early mesodermal marker (known as *brachyury* in other vertebrates), was not activated in the ring of cells at the embryonic margin.

During normal embryogenesis, initial expression of *ntl* is associated with a specific deformation in the shape of the cells lying at the border of the growing embryo. This margin is, however, not fixed: it advances as gastrulation proceeds, and its cellular composition changes as cells transit through the margin before

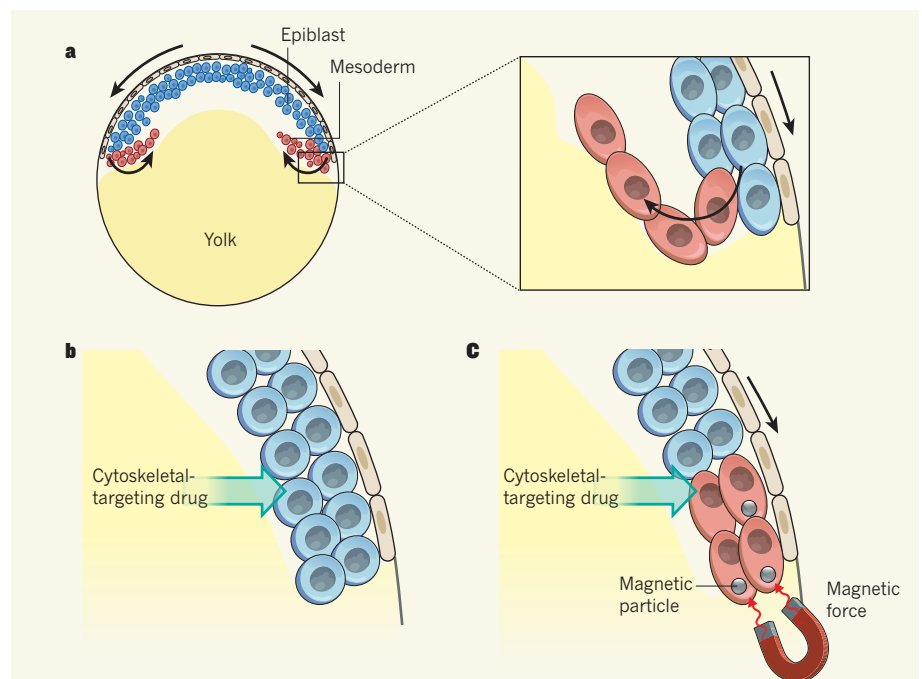


Figure 1 | Stretched embryos make mesoderm. During embryonic development, a process called gastrulation leads to the formation of three cell layers — the ectoderm, mesoderm and endoderm. **a**, In zebrafish, gastrulation involves a mass movement, or epiboly, of a thick layer of cells called the epiblast. During epiboly, cells at the border of the developing embryo undergo structural deformation. **b**, Brunet *et al.*⁴ show that, in embryos treated with drugs that block the function of the cytoskeleton (and so block cellular deformation), epiboly and early mesoderm formation are inhibited. **c**, However, they find that external application of strain — by injecting the embryonic cells with magnetic particles and then literally pulling on them with a magnet — rescues epiboly movements and mesoderm-specific gene expression.

rolling inside the embryo. Might mesoderm gene expression be activated by structural cell deformation associated with this border? To answer this question, Brunet *et al.* injected the mechanically defective embryos with magnetic particles, allowing the margin cells to be mechanically towed by an electromagnet ring positioned around the embryo. Remarkably, this procedure rescued cell deformation, epiboly movements and mesoderm gene expression (Fig. 1c). Furthermore, rescue occurred even in neighbouring cells that did not receive the magnetic particles, suggesting that mechanical forces are transmitted through cell–cell junctions.

Brunet *et al.* next studied *Drosophila* (fruit-fly) embryos lacking the *snail* gene — these also show impaired mesoderm formation⁶. Again, they found that a mechanical challenge, in the form of external application of an indentation, restored mesoderm-formation competence in the embryos, suggesting that the relevance of mechanical interactions to this process is evolutionarily conserved.

Although Brunet and colleagues did not study the gastrulation mechanics of mammals and other amniotes, it is worth mentioning that in mouse and chick embryos epiboly seems to be uncoupled from mesoderm formation; instead, the latter occurs at a structure known as the primitive streak, which lies in the middle of the embryo¹. However, at least in chick embryos, formation of the primitive streak is preceded by massive cell–cell intercalations, analogous to the epiboly movements in fish embryos. Intriguingly, interfering with these cellular movements in chick embryos does not block mesoderm formation, but causes it to form at the embryo boundary⁷, as in fish and other anamniotes. Thus, one might speculate that mesoderm induction in amniotes also requires an association with a mechanically challenged cellular state, as Brunet *et al.* predict. In line with this idea, it was recently shown⁸ that, in mouse embryos, mechanical forces from maternal tissues have a causal role in forming the specialized tissue responsible for inducing the anterior–posterior axis.

So it seems that, after years of oblivion, research on mechanical forces is returning to centre stage in cell biology, promising to enhance our understanding of the links between cell behaviour and the structural context of cells and organs⁹. Such studies will also further our understanding of mechanotransduction pathways, which link physical forces to cell signalling. Brunet *et al.* contribute to this, by showing that the shuttling of the signalling molecule β -catenin from the cytoplasm to the nucleus can be used as a proxy for mechanical activation of border cells in the early phases of gastrulation in zebrafish and fly embryos. It seems that this nuclear shuttling occurs independently of the Wnt signalling pathway (with

which β -catenin is typically associated), but instead involves phosphorylation of β -catenin by Src kinase enzymes. Clearly, further work is required to confirm the relevance of β -catenin as a mechanoeffector in gastrulation, given that some of the authors' claims about the function of this protein rest on experiments using overexpression of dominant-negative proteins.

Brunet and colleagues' work reminds us that cells can extract as much information from changes in their form or three-dimensional tissue localization as they do from soluble factors. How the genome responds to these changes, and whether the embryo uses mechanical forces as a long-range communication device to orchestrate its own morphogenesis, remain questions for future studies. ■

CANCER

A suppression switch

The status of the protein p53 determines whether inhibiting the cellular autophagy pathway promotes or inhibits pancreatic cancer in mice. This finding serves as a cautionary tale for clinical trials of autophagy inhibitors. [SEE LETTER P.296](#)

HANNA STAROBINETS & JAYANTA DEBNATH

Autophagy is a fundamental process in which a cell cannibalizes itself, degrading and recycling cytoplasmic proteins and organelles. This pathway has a crucial role in promoting cellular homeostasis and survival in response to diverse forms of stress, so there is considerable interest in modulating autophagy in cancer cells. So far, clinical trials have focused predominantly on enhancing the efficacy of chemotherapy by inhibiting autophagy, using antimalarial drugs such as hydroxychloroquine¹. But there is confusion as to whether inhibiting autophagy enhances or diminishes cancer therapy — current evidence suggests that both may be true. In this issue, Rosenfeldt *et al.*² (page 296) identify the tumour-suppressor protein p53 as a determinant of whether autophagy suppresses or accelerates the progression of pancreatic cancer.

Pancreatic cancers, specifically pancreatic ductal adenocarcinomas (PDACs), are aggressive and lethal tumours that commonly display mutational activation of the signalling molecule Kras³. Recent work has established that tumours characterized by mutations in Kras or other Ras proteins rely on autophagy for growth and cell proliferation, making this pathway an attractive therapeutic target^{4–7}. Rosenfeldt and colleagues used genetically engineered mice to investigate the role of autophagy in the progression and treatment of PDACs driven by the mutation Kras^{G12D}. They demonstrate that silencing essential

Stefano Piccolo is in the Department of Molecular Medicine, University of Padua Medical School, 35126 Padua, Italy.
e-mail: piccolo@bio.unipd.it

1. Stern, C. D. *Gastrulation: From Cells to Embryo* (Cold Spring Harb. Lab. Press, 2004).
2. De Robertis, E. M. *Cell* **132**, 185–195 (2008).
3. Technau, U. & Scholz, C. B. *Int. J. Dev. Biol.* **47**, 531–539 (2003).
4. Brunet, T. *et al. Nature Commun.* <http://dx.doi.org/10.1038/ncomms3821> (2013).
5. Wozniak, M. A. & Chen, C. S. *Nature Rev. Mol. Cell Biol.* **10**, 34–43 (2009).
6. Desprat, N., Supatto, W. & Pouille, P.-A. *Dev. Cell* **15**, 470–477 (2008).
7. Voiculescu, O., Bertocchini, F., Wolpert, L., Keller, R. E. & Stern, C. D. *Nature* **449**, 1049–1052 (2007).
8. Hiramatsu, R. *et al. Dev. Cell* **27**, 131–144 (2013).
9. Halder, G., Dupont, S. & Piccolo, S. *Nature Rev. Mol. Cell Biol.* **13**, 591–600 (2012).

autophagy-regulating proteins (either ATG5 or ATG7) in Kras^{G12D}-mutant pancreatic epithelial cells led to higher expression of p53, which was accompanied by decreased proliferation, increased apoptotic cell death and elevated cellular senescence, all of which are important barriers to tumour formation⁸.

The authors further show that this loss of autophagy is sufficient to prevent the progression of early stage precancerous lesions, termed pancreatic intraepithelial neoplasias (PanINs), into more advanced cancers (Fig. 1a). This finding is consistent with previous work demonstrating a requirement for autophagy in the growth of pancreatic cancer⁹. Interestingly, Rosenfeldt *et al.* also show that the engineered loss of autophagy in normal mouse pancreatic tissue led to elevated p53 expression and cell death, and that this resulted in pancreatic-tissue destruction and diabetes.

More than half of human PDACs exhibit silencing or mutation of the gene encoding p53 (ref. 10), raising the question of whether defective autophagy will still prevent PDAC progression when p53 is inactivated. The authors tested the effects of combined autophagy loss and p53 deficiency in Kras-mutant PDACs and, surprisingly, found that this accelerated, rather than impeded, PDAC progression (Fig. 1b). In a key experiment, the authors treated mice that had Kras^{G12D}-driven, p53-deficient lesions with hydroxychloroquine and again observed significantly faster PDAC formation. This result contrasts with the previous observation⁹ of delayed tumour

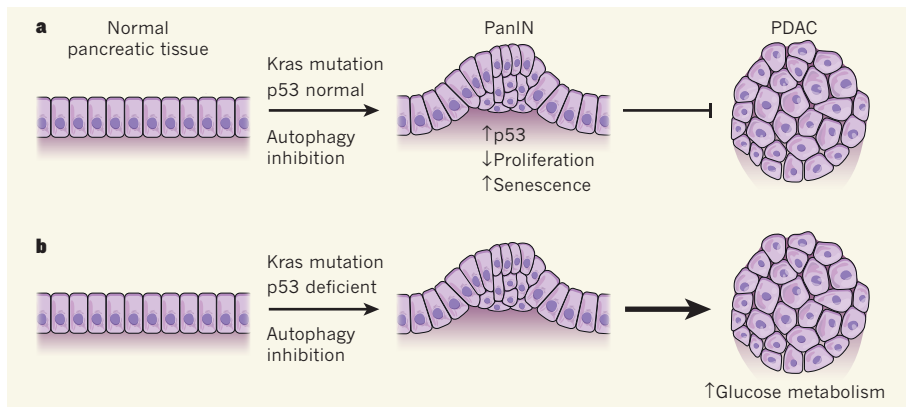


Figure 1 | Autophagy, p53 and cancer progression. Mutations that cause abnormal activation of the protein Kras are commonly associated with pancreatic cancer. The cell proliferation that results from these mutations leads to the development of precancerous lesions called pancreatic intraepithelial neoplasias (PanINs), which stochastically develop into invasive pancreatic ductal adenocarcinomas (PDACs). **a**, Rosenfeldt *et al.*² show that when this process is accompanied by normal activity of the tumour-suppressor protein p53, inhibiting autophagy blocks tumour progression at the PanIN stage, which is associated with p53 activation, suppression of proliferation and increased cellular senescence. **b**, However, if the Kras-mutant pancreatic cells lack p53, inhibition of autophagy accelerates the development of PDACs. The authors suggest that this acceleration may be due to enhanced glucose metabolism.

progression following treatment of Kras^{G12D}-driven, p53-normal PDACs with chloroquine, a hydroxychloroquine derivative.

Thus it seems that p53 acts as a switch in pancreatic cancer that dictates whether therapeutic inhibition of autophagy slows or accelerates disease progression. It remains unclear whether p53 similarly regulates autophagy inhibition in other cancers, but it seems likely that there will be cancer-specific nuances — in Kras-mutant lung cancers, for example, silencing of ATG7 suppresses proliferation and alters tumour differentiation, irrespective of p53 status⁶. Nevertheless, Rosenfeldt and colleagues' findings have immense clinical implications, because they highlight the importance of determining the p53 status of pancreatic cancers before treatment with autophagy inhibitors.

The activation of Ras proteins elicits profound metabolic changes that drive energy production and biosynthetic capacity in rapidly proliferating tumour cells; previous studies^{4–7} have demonstrated a requirement for autophagy in sustaining cellular metabolism during Ras mutation. In mouse models of lung cancers, autophagy-deficient precancerous tumours harbouring mutations in Kras or Braf (another signalling molecule commonly mutated in cancer) are unable to progress to the malignant stage and exhibit impaired mitochondrial metabolism^{5,6}. By contrast, Rosenfeldt *et al.* propose that increased glucose metabolism is responsible for the accelerated progression of Kras^{G12D}-driven PDACs seen following concomitant inhibition of p53 and autophagy.

In support of this, the authors show that cells from PDACs growing in Kras^{G12D} mice that also lacked p53 and ATG7 exhibited enhanced glucose uptake and increased metabolite levels

compared with their autophagy-proficient Kras^{G12D}, p53-lacking counterparts. However, restoration of autophagy by re-expression of ATG7 did not reverse these metabolic alterations, so it remains unclear whether the metabolic changes are a cause or a consequence of the increased aggressiveness displayed by PDAC cells lacking both autophagy and p53. Furthermore, the loss of autophagy may have other metabolic consequences: in Kras^{G12D}-driven lung cancers, for example, the combined loss of ATG7 and p53 results in aberrant fatty-acid oxidation and profound lipid accumulation, suggesting a role

for autophagy in lipid breakdown^{5,6}.

Finally, it is important to recognize that in human Kras-mutant PDAC cell lines with p53 mutations, the loss of autophagy reduces proliferation and tumour growth⁹ — the opposite effects to those described by Rosenfeldt *et al.* in mice. This discrepancy may arise from the different effects of p53 mutation versus outright genetic deletion on metabolism in pancreatic cancers. Although further study is required to understand the mechanism underlying the ability of p53 to switch the clinical outcome of autophagy inhibition, Rosenfeldt and colleagues have illustrated the importance of defining the molecular contexts in which targeting autophagy may be beneficial for anticancer therapy. ■

Hanna Starobinets and Jayanta Debnath are in the Department of Pathology and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco 94143, USA.
e-mail: jayanta.debnath@ucsf.edu

1. Yang, Z. J., Chee, C. E., Huang, S. & Sinicrope, F. A. *Mol. Cancer Therapeut.* **10**, 1533–1541 (2011).
2. Rosenfeldt, M. T. *et al. Nature* **504**, 296–300 (2013).
3. Morris, J. P., Wang, S. C. & Hebrock, M. *Nature Rev. Cancer* **10**, 683–695 (2010).
4. Guo, J. Y. *et al. Genes Dev.* **25**, 460–470 (2011).
5. Guo, J. Y. *et al. Genes Dev.* **27**, 1447–1461 (2013).
6. Strohecker, A. M. *et al. Cancer Discov.* **3**, 1272–1285 (2013).
7. Lock, R. *et al. Mol. Biol. Cell* **22**, 165–178 (2011).
8. Lowe, S. W., Cepero, E. & Evan, G. *Nature* **432**, 307–315 (2004).
9. Yang, S. *et al. Genes Dev.* **25**, 717–729 (2011).
10. Rozenblum, E. *et al. Cancer Res.* **57**, 1731–1734 (1997).

This article was published online on 4 December 2013.

ASTROPHYSICS

Tracking our neighbours' past

The collective motions of dwarf galaxies in planes around the Andromeda galaxy and the Milky Way have presented a challenge to theory. Interactions between galaxy groups in the distant past may have left their imprint on these dwarfs.

ALAN MCCONNACHIE

Little galaxies have long been known to orbit around large galaxies such as our own Milky Way or our neighbour Andromeda (Fig. 1). But recent discoveries^{1,2} showing that the nearest 'dwarf' galaxies are collectively rotating in giant planar structures — akin to disks — have baffled astronomers. Writing in *Monthly Notices of the Royal Astronomical Society*, Shaya and Tully³ describe an

analysis suggesting that the dwarfs' peculiar motion is an imprint of a time when Andromeda and the Milky Way were not the only big kids on the block.

The same theory that has shown so much success in predicting the large-scale properties of the Universe — for example, the cosmic microwave background radiation that permeates the whole of space, detected in fine detail by the Planck satellite earlier this year⁴ — also makes predictions about how dwarf galaxies

form. The expectation is that they do so essentially independently of each other. So it came as a surprise when it was reported in January¹ that many of the dwarfs around Andromeda move as a group, mapping out a giant, rotating, thin disk that is viewed nearly edge on from Earth. This arrangement seems to be similar to that shown by the Milky Way's dwarfs².

How can a bunch of objects that form separately conspire to share a property as fundamental as how they move around their larger hosts? If the theory cannot explain the data, then the theory must be wrong. Astronomically speaking, the stakes are high, and an important first step is to work out whether there is a way — any way — in which the standard cosmological theory could match the seemingly bizarre movements of our littlest neighbours.

To this end, Shaya and Tully attempt to reconstruct the orbits of all our Galactic neighbours from the early ages of the Universe up until now, to see whether they can identify how and when these disks of dwarfs came to be. The authors find many possible solutions, but it is when they dig into the details of these solutions that the elegance of their approach starts to shine through. As a result of cosmic expansion, many galaxies that were once relatively close to us are now much farther away, and the researchers isolate key moments, and key protagonists, that they identify as shouldering the responsibility for the formation of the planes of dwarfs.

The most significant players seem to be two vast features: the Local Sheet and the Local Void. The Local Sheet is a giant structure that contains large numbers of galaxies and galaxy groups — including our own — and forms a wall along the edge of a vast expanse of emptiness, known as the Local Void. Shaya and Tully suggest that, in the early Universe, Andromeda was pulled out of this void, dragging behind it an entourage of dwarf galaxies that were attracted to Andromeda by its gravitational pull. Andromeda, being attracted to the Milky Way, effectively stretched these dwarfs out into a chain aligned with the Milky Way, and it is this feature that we now observe as the disk of dwarfs around Andromeda.

Meanwhile, the Milky Way was caught up in its own struggle for companions with Andromeda and another galaxy group, now much farther away. Because of their competing gravitational pulls, only certain dwarfs were pulled towards the Milky Way. And all this proceeded as the Milky Way was pulled towards the Virgo Cluster — a massive collection of galaxies now nearly 20 times farther away than Andromeda. Aiming at a moving target, the dwarfs that chased the Milky Way now find themselves aligning in the direction of this massive, but now distant, structure.

The spatial distribution and motion of dwarf galaxies are just the latest in a series of problems that have cropped up as researchers



Figure 1 | The Andromeda galaxy.

explore the consequences of the favoured cosmological model in ever more detail on scales for which the theory was not originally developed. The number of dwarfs^{5,6}, the mass distribution within dwarfs^{7,8} and now the spatial distribution of dwarfs are all new testing grounds for cosmology that are providing challenges to our understanding, particularly when considered alongside large-scale phenomena such as galaxy clustering or the cosmic microwave background.

So, has the model for how the Universe puts itself together survived another observation? Maybe. What is offered by Shaya and Tully is not a definitive answer — far from it, as they are at pains to stress. Their method is designed to find solutions, irrespective of whether such solutions are 'likely'. What they do show is that, with the right set of interactions, weird-looking arrangements of dwarfs like those we observe around the Milky Way and Andromeda might be explained without any need to change anything fundamental in the standard model — which is not the same as saying that changing something fundamental might not be better.

But Shaya and Tully's analysis stands out because it provides invaluable insight into a

complex arena in which galaxies now separated by vast distances compete for companions, and dwarfs chase large galaxies through space as their large hosts are pulled towards distant clusters. It is an elegant demonstration that the complexity of the structure of the Universe at large can imprint itself on the smallest scales. And, even in astronomy, what seem to be local problems may well necessitate global solutions. ■

Alan McConnachie is at the NRC Herzberg Institute of Astrophysics, Dominion Astrophysical Observatory, Victoria, British Columbia V9E 2E7, Canada. e-mail: alan.mcconnachie@nrc-cnrc.gc.ca

1. Ibata, R. A. *et al.* *Nature* **493**, 62–65 (2013).
2. Kroupa, P., Theis, C. & Boily, C. M. *Astron. Astrophys.* **431**, 517–521 (2005).
3. Shaya, E. J. & Tully, R. B. *Mon. Not. R. Astron. Soc.* **436**, 2096–2119 (2013).
4. Planck Collaboration *Astron. Astrophys.* Preprint at <http://arxiv.org/abs/1303.5062> (2013).
5. Moore, B. *et al.* *Astrophys. J.* **524**, L19–L22 (1999).
6. Klypin, A. *et al.* *Astrophys. J.* **522**, 82–92 (1999).
7. de Blok, W. J. G. *et al.* *Astron. J.* **136**, 2648–2719 (2008).
8. Boylan-Kolchin, M., Bullock, J. S. & Kaplinghat, M. *Mon. Not. R. Astron. Soc.* **415**, L40–L44 (2011).

NEUROSCIENCE

The highs and lows of memory

Recent experience can affect memory either positively or negatively. A small population of inhibitory neurons that express the protein parvalbumin seems to have a pivotal role in this process. [SEE LETTER P.272](#)

KEVIN ALLEN & HANNAH MONYER

The ability to memorize our daily life depends on a brain region called the hippocampus. Patients with damage to this region are unable to form new long-term memories of day-to-day events¹. Given its importance, the hippocampus has been the subject of a tremendous amount of research aimed at understanding how the hippocampal neuronal network supports memory. Writing on page 272 of this issue, Donato and colleagues² provide insight into how recent experience can alter hippocampal memory functions. The main contribution of this work is the identification of bidirectional, experience-dependent modifications in a subset of hippocampal neurons that are associated with enhanced or impaired memory performance.

Neurons in the hippocampal network can be divided into two main categories: excitatory and inhibitory. Excitatory neurons are the main carriers of information and connect the hippocampal network with other brain

regions. But their activity is not left unchecked: excitatory neurons are tightly regulated by the release of the neurotransmitter GABA from inhibitory neurons. In fact, it has been suggested^{3,4} that many of the extraordinary computational processes implemented in neuronal networks originate from the action and diversity of inhibitory neurons.

Donato *et al.* investigated the effect of previous experience on a group of inhibitory neurons that express the calcium-buffering protein parvalbumin. Although this group constitutes only some 5% of the hippocampal neuronal population, it is crucial for sustaining several types of rhythmic neural activity in the brain and has been implicated in numerous cognitive processes⁵.

The authors assigned mice to one of two experimental conditions. One group of animals was housed in an enriched environment rather than the standard conditions common to most laboratories. The second group was subjected to contextual fear conditioning, which involves placing the animals in an environment where

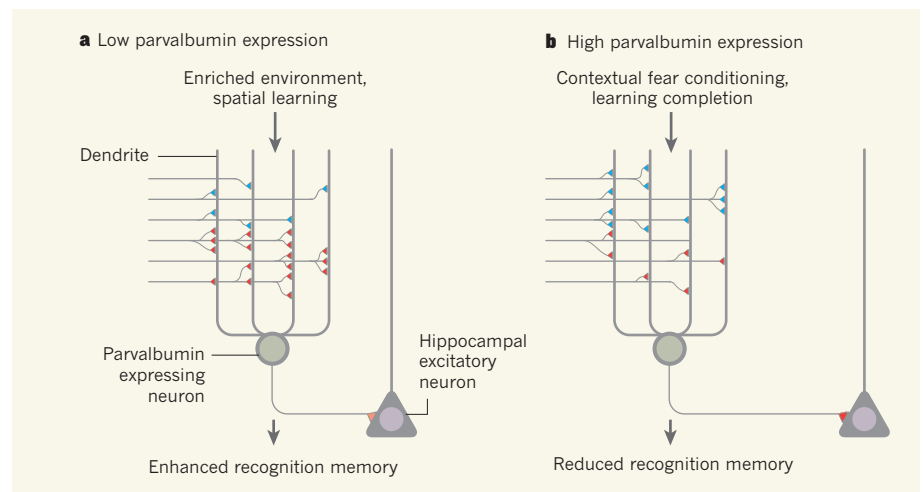


Figure 1 | Experience and alteration of parvalbumin-expressing neurons. **a**, Donato *et al.*² propose that an enriched environment and spatial learning can increase the number of inhibitory synaptic terminals (red) that make contact with parvalbumin-expressing neurons, leading to low parvalbumin expression. This affects hippocampal excitatory neurons downstream, resulting in enhanced memory. Excitatory synaptic terminals are shown in blue. The pink synaptic terminal is inhibitory, and expresses low levels of parvalbumin and the enzyme GAD67. **b**, By contrast, contextual fear conditioning and completion of learning increase the number of excitatory synaptic terminals that make contact with parvalbumin-expressing neurons, which is associated with high parvalbumin expression and, consequently, reduced recognition memory at the level of hippocampal neurons.

they receive a series of five electric shocks.

The team found that the enriched environment resulted in a decrease in parvalbumin expression, whereas contextual fear conditioning caused the opposite (Fig. 1). Intriguingly, parvalbumin expression was strongly correlated with that of the enzyme GAD67, which in inhibitory neurons transforms the amino acid glutamate into GABA. This effect of recent experience on parvalbumin levels seemed to be restricted to basket cells, a subgroup of parvalbumin-expressing neurons that makes synaptic connections predominantly onto the cell body of excitatory and other inhibitory neurons. Donato and co-authors propose that the change in parvalbumin expression resulted from a shift in the ratio of excitatory to inhibitory synaptic inputs onto basket cells.

The most surprising finding of this work is that experience-dependent changes in parvalbumin expression coincide with changes in memory functions. Mice with low parvalbumin expression following exposure to an enriched environment showed enhanced object-recognition memory, whereas those with high parvalbumin expression after fear conditioning displayed impaired memory. To gather further evidence of a role for parvalbumin expression level in memory, the authors used a pharmacogenetic approach to manipulate the activity of cells expressing parvalbumin. Inhibiting or activating parvalbumin-expressing neurons led to a decrease or increase in parvalbumin expression levels, respectively. These manipulations were associated with the expected behavioural outcome, such that experimentally controlled parvalbumin expression was negatively correlated with object-recognition memory.

Donato *et al.* also assessed parvalbumin expression levels during the course of spatial learning. They found that, in a navigation task, parvalbumin expression was low during the learning phase of the task, but shifted to high once learning was completed. Again, low and high parvalbumin expression were associated with enhanced and reduced object-recognition memory, respectively. These learning-related changes in parvalbumin expression were also accompanied by a modification of the synaptic inputs onto basket cells.

One of the most pressing questions in light of this work is how neuronal activity in the hippocampal network changes with variations in parvalbumin expression. Future studies should address the consequences of enriched environments and contextual fear conditioning on the activity of parvalbumin-expressing neurons within the hippocampal network. The activation of parvalbumin-expressing basket cells during high and low expression of parvalbumin could also be investigated in brain sections by measuring the response of these neurons after stimulation of input pathways. Given the anatomical synaptic alterations presented in this study, one expects modifications of the

excitatory and inhibitory synaptic currents during high and low expression states of parvalbumin, respectively. Moreover, changes in the calcium-buffering capacity and in GAD67 levels should affect the synaptic output of parvalbumin-expressing neurons^{6,7}.

A second avenue that deserves further investigation relates to the link between recent experience and parvalbumin expression. Donato and co-workers suggest that the balance of excitatory and inhibitory synapses onto parvalbumin neurons is a key factor governing the expression of this protein. But what determines these changes in synaptic connectivity in the first place? Are they the result of specific activity patterns occurring in the hippocampus or are they instead controlled by neuromodulatory signals originating outside the hippocampus? Answers to these questions would complete the picture of how recent

experience modulates learning and memory by altering the configuration of the parvalbumin-expressing neuronal network. ■

Kevin Allen and Hannah Monyer are in the Department of Clinical Neurobiology, Medical Faculty of Heidelberg University and German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.
e-mail: h.monyer@dkfz-heidelberg.de

1. Scoville, W. B. & Milner, B. J. *Neuropsychiatry Clin. Neurosci.* **12**, 103–113 (2000).
2. Donato, F., Rompani, S. B. & Caroni, P. *Nature* **504**, 272–276 (2013).
3. Klausberger, T. *et al. Nature* **421**, 844–848 (2003).
4. Freund, T. F. & Buzsáki, G. *Hippocampus* **6**, 347–470 (1996).
5. Fuchs, E. C. *et al. Neuron* **53**, 591–604 (2007).
6. Obata, K. *et al. Biochem. Biophys. Res. Commun.* **370**, 429–433 (2008).
7. Eggermann, E. & Jonas, P. *Nature Neurosci.* **15**, 20–22 (2011).

SOFT-MATTER PHYSICS

Ferromagnetic ferrofluids

The idea of creating liquids containing magnetic particles that display ferromagnetism has not been realized — until now. Plate-shaped magnets in a liquid crystal have been made that exhibit this property. [SEE ARTICLE P.237](#)

NOEL A. CLARK

At room temperature, iron atoms behave as identical sub-nanometre-sized bar magnets. Solid iron's crystal structure is such that if a piece of the metal — say, a nail — is placed in a magnetic field, the atomic magnets, which will tend to align along the

field, will maintain their induced orientation even when the field is removed. The nail itself will thus become a big bar magnet (Fig. 1a). As a result, materials that hold magnetic ordering in the absence of an applied magnetic field are named after iron and termed ferromagnetic, or just magnetic for short¹. On page 237 of this issue, Mertelj *et al.*² report the observation of

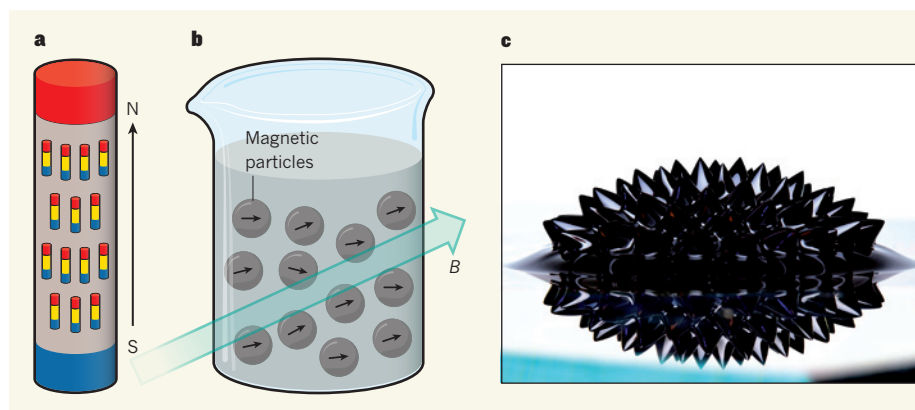


Figure 1 | Bar magnets and ferrofluids. **a**, In ferromagnetic iron, the magnetic moments of iron atoms, here illustrated as tiny bar magnets, maintain their mutual south–north orientation in the absence of a magnetic field to form a bigger, permanent bar magnet. Its magnetic moment is shown by the arrow. **b**, Ferrofluids, suspensions of sub-micrometre-sized magnetic particles, respond strongly to an applied magnetic field (B), forcing the fluid to form dramatic spikes (**c**) as a way of filling its volume with field. **c**, Photograph of a ferrofluid on a reflective glass plate and subjected to a strong magnetic field.

C. GREGORY F. MAXWELL/WIKIMEDIA

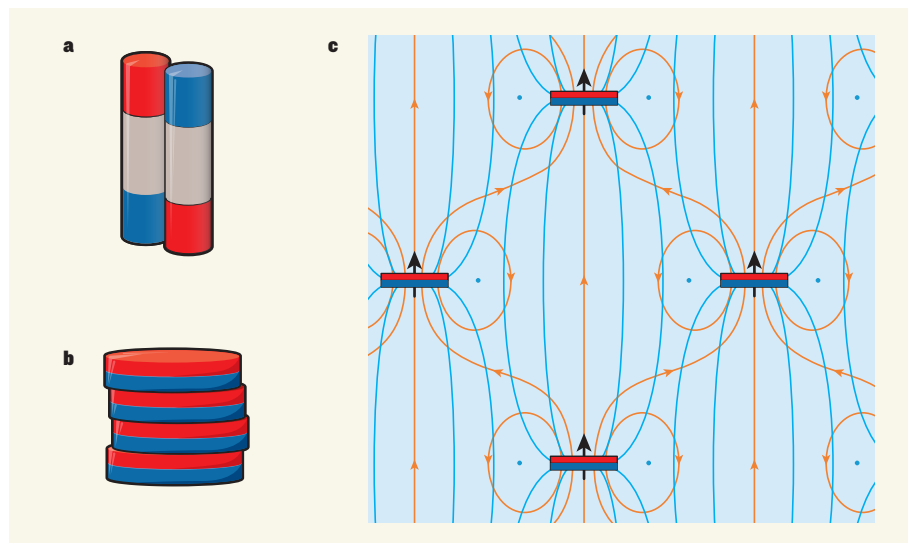


Figure 2 | Ferromagnetic order in a ferrofluid. **a**, Ferrofluids are non-magnetic in the absence of a magnetic field because of thermal disorder and the fact that their constituent micro-magnets arrange in pairs that cancel out the net field that the magnets produce. **b**, Mertelj *et al.*² used plate-shaped magnets with moments that add upon pairing. **c**, The authors' ferromagnetic ferrofluid uses a nematic liquid crystal (blue) with local molecular orientation (blue lines) in order to keep the magnetic plates horizontal in the image. With this constraint, the magnetic moments of the plates (black arrows) mutually orient parallel to one another, as in the stacking of magnetic plates in **b** and as can be seen from the continuity of the magnetic-field lines (orange).

ferromagnetism in a novel marriage of particle and fluid.

One of the most beautiful applications of ferromagnets is ferrofluids. These are suspensions of sub-micrometre-sized magnetic particles in organic solvents, such as kerosene, with a surfactant soap to keep them dispersed³. The micro-magnets in a ferrofluid also respond to a magnetic field by orienting along the field (Fig. 1b) and, as a result, the field can exert strong forces on the fluid, radically changing its shape as the fluid moves to maximally fill its volume with magnetic field (Fig. 1c). These effects are both exotic and useful, with a wide range of applications, ranging from computer disk drives to rotary vacuum seals and speaker-damping technology.

However, an important feature of ferrofluids is that, despite their dramatic response to fields, they are not macroscopically ferromagnetic. That is, although their micro-particles are ferromagnetic, once an applied field is removed the induced magnetic ordering averaged over volumes containing many particles disappears. Such a material — with no remnant magnetization in the absence of field — is termed paramagnetic. When iron is heated to temperatures above 1,050 kelvin, at which point thermal disordering of its atomic magnetic-moment orientations dominates, it also becomes paramagnetic¹.

Ferrofluids are paramagnetic because of thermal disorder of particle orientation at room temperature and because there is a tendency for magnetic particles in suspension to arrange in pairs or groups that cancel out the net field that the particles produce. Anyone

who has played with bar magnets is familiar with the way they snap together (Fig. 2a) with their magnetic dipoles oriented in opposite directions, mutually cancelling out each other's moment. However, the example of iron shows that ferromagnetism is interesting and useful, motivating the question of whether ferromagnetic ferrofluids could be made, and how their properties would differ from those of currently known paramagnetic ferrofluids.

Mertelj *et al.* take a major step towards addressing this question, combining techniques of soft-matter and magnetic-particle physics to fundamentally change each of the key ferrofluid components (particles, solvent and soap) to create a system in which the field-cancelling magnetic micro-particle association is suppressed to yield stable ferromagnetic order in a fluid suspension (Fig. 2c).

To appreciate their achievement further, more playing with magnets is in order — this time, with plate-like magnets in which the magnetic moment is normal to the plane of the plate (Fig. 2b). This reveals that field cancellation is more difficult with plate-shaped magnets than it is for rod-shaped ones. The plates tend to stack strongly like coins, their fields adding, but form only weak side-by-side antiparallel geometries, an arrangement more likely to yield ferromagnetic ordering in suspension. So Mertelj *et al.* used micro-plates of barium hexaferrite — a magnetic material that can be grown in ultrathin crystalline plates of high aspect ratio⁴ (5 nanometres thick and, on average, 70 × 70 nm in area) — that have their atomic magnetic moments normal to the plane of the plate.

To make a suspension of these magnetic plates, Mertelj and colleagues chose as the solvent a nematic liquid crystal (LC; Fig. 2c), a non-magnetic fluid of nanometre-scale molecular rods that spontaneously acquire long-range orientational order — a propensity for the molecular orientation to become as uniform as possible given the constraints of the physical boundaries of the fluid's container⁵. For the present purposes, the LC is simply a host, but one with internal orientational elasticity.

On to the surfactant, which Mertelj *et al.* use to couple the orientations of the particles and the LC, coating the barium hexaferrite plates so that the LC orientation at the plates' surfaces is normal to the plates. This, in turn, causes the plates to align normal to the average LC direction, keeps them flat and gives them a charge, causing them to repel one another — a combination that enables them to fill space as a fog of parallel flat flakes (Fig. 2c). Now, in principle, some flakes could have their magnetic moments up and others down. But, as in the game with plates, the tendency for uniform orientation will win out and large volumes of uniform magnetic orientation will spontaneously grow to form ferromagnetic domains in a ferrofluid.

In addition to creating a ferromagnetic ferrofluid, Mertelj and colleagues have made a LC system that is ferromagnetic on length scales that incorporate many particles, the first LC system to exhibit orientational ordering that changes in a polar manner in response to an applied magnetic field. This discovery opens up the possibility of producing a wide variety of new magnetic-field effects in which LC molecular orientation, and therefore its influence on light propagation, can be controlled by an applied magnetic field. It also suggests that it might be feasible to make electrical analogues — similarly designed suspensions of electrically polar particles. However, the understanding and application of field-induced manipulation of LC molecular orientation, in which electric fields have enabled LC displays and the portable computing revolution, seems to be heading for a magnetic revolution based on ferromagnetic coupling of optical response and magnetic field. ■

Noel A. Clark is in the Department of Physics, University of Colorado, Boulder, Colorado 80309-0390, USA.
e-mail: noel.clark@colorado.edu

1. Cullity, B. D. & Graham, C. D. *Introduction to Magnetic Materials* (Wiley, 2009).
2. Mertelj, A., Lisjak, D. & Drogenik, M. & Čopič, M. *Nature* **504**, 237–241 (2013).
3. Odenbach, S. *Ferrofluids: Magnetically Controllable Fluids and Their Applications* (Springer, 2002).
4. Ovtar, S., Lisjak, D. & Drogenik, M. *J. Colloid Interface Sci.* **337**, 456–463 (2009).
5. Oswald, P. & Pieranski, P. *Nematic and Cholesteric Liquid Crystals: Concepts and Physical Properties Illustrated by Experiments* (CRC Press, 2005).

An archaeal origin of eukaryotes supports only two primary domains of life

Tom A. Williams¹, Peter G. Foster², Cymon J. Cox³ & T. Martin Embley¹

The discovery of the Archaea and the proposal of the three-domains ‘universal’ tree, based on ribosomal RNA and core genes mainly involved in protein translation, catalysed new ideas for cellular evolution and eukaryotic origins. However, accumulating evidence suggests that the three-domains tree may be incorrect: evolutionary trees made using newer methods place eukaryotic core genes within the Archaea, supporting hypotheses in which an archaeon participated in eukaryotic origins by founding the host lineage for the mitochondrial endosymbiont. These results provide support for only two primary domains of life—Archaea and Bacteria—because eukaryotes arose through partnership between them.

Since their discovery by Carl Woese and his co-workers in 1977, the Archaea have figured prominently in hypotheses for eukaryotic origins^{1,2}. Although similar to Bacteria in terms of cell structure, molecular phylogenies for ribosomal RNA and a small core of genes, that mainly have essential roles in protein translation³, suggested that the Archaea were more closely related to the eukaryotic nuclear lineage; that is, to the host cell that acquired the mitochondrion⁴. The idea that Archaea and eukaryotes are more closely related to each other than either is to Bacteria depends on analyses suggesting that the root of the tree should be placed on the bacterial stem, or within the Bacteria^{5–12}, implying that the prokaryotes—cells that lack a nucleus—are a paraphyletic group¹³. The main question now debated is whether core components of the eukaryotic nuclear lineage descend from a common ancestor shared with Archaea, as in the three-domains tree¹⁴ (Fig. 1), which is also often called the ‘universal tree’ or ‘tree of life’^{15–17}, or from within the Archaea, as proposed by archaeal-host hypotheses for eukaryotic origins². The archaeal-host scenario with the greatest phylogenetic support is the eocyte hypothesis¹⁸, which proposes a sister-group relationship between eukaryotes and the

eocytes (or Crenarchaeota¹⁴), one of the major archaeal divisions (Fig. 1). However, the three-domains–eocyte debate remains controversial because different phylogenetic methods have delivered different results, often from the same data¹⁹. This disagreement is due, at least in part, to the difficulties associated with resolving ancient divergences in phylogenetic trees.

Challenges of reconstructing ancient relationships

A major issue in reconstructing ancient relationships is the strength and quality of historical signal remaining after the millions of years since the divergence of Archaea and eukaryotes. The earliest fossils identified as eukaryotic appeared by about 1.8 billion years ago²⁰; over this enormous span of time, the accumulation of multiple substitutions in DNA and protein sequences might have erased any signal that would allow the relationship between archaeal and eukaryotic core genes to be established²¹. However, more recent simulations and empirical studies suggest that there are reasons to be cautiously optimistic that this is not the case: functional constraints vary across real DNA and protein sequences so that sites evolve at different rates^{22–25}. Fast-evolving sites are indeed

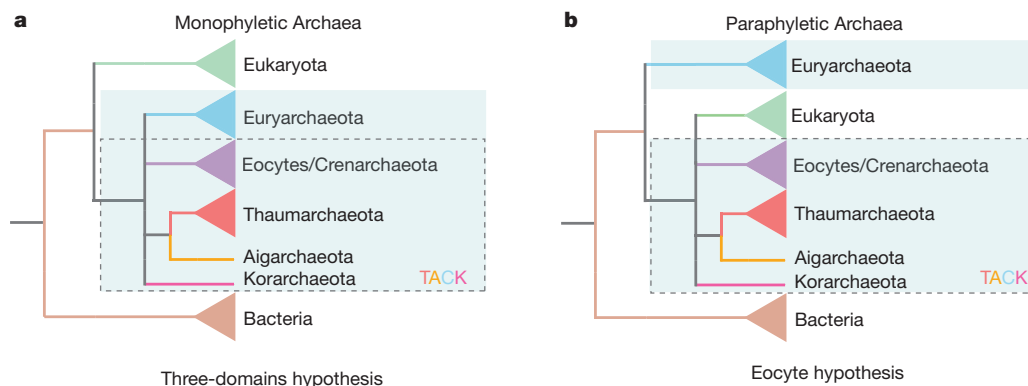


Figure 1 | Competing hypotheses for the origin of the eukaryotic host cell. **a**, The rooted three-domains tree¹⁴ depicts cellular life divided into three major monophyletic groups or domains: the Bacteria, Archaea and Eukaryota—the latter representing the host lineage, sometimes also called the nuclear or nucleo-cytoplasmic lineage⁵, that acquired the mitochondrial endosymbiont. In this tree the Archaea and Eukaryota are most closely related to each other because they share a common ancestor that is not shared with Bacteria. **b**, The rooted eocyte tree recovers the host-cell lineage nested within the

Archaea as a sister group to the eocytes (which Woese *et al.*¹⁴ called the Crenarchaeota); this implies that, on the basis of the small set of core genes, there are only two primary domains of life—the Bacteria and the Archaea. In its modern formulation shown here the eocyte hypothesis implies that the closest relative of the eukaryotic nuclear lineage is one, or all, of the TACK Archaea, which include newly discovered relatives of the eocytes/Crenarchaeota. Both trees have been traditionally rooted on the bacterial stem, consistent with some published analyses^{5–8}.

¹Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne NE2 4HH, UK. ²Department of Life Sciences, Natural History Museum, London SW7 5BD, UK. ³Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

quickly saturated but the slowest sites can still retain useful phylogenetic information, explaining why we are able to align some genes over the entire tree of life. Analyses of molecular sequences might therefore be able to distinguish between the alternative hypotheses for eukaryotic core gene origins, but the phylogenetic methods used and the types of data analysed are likely to be of critical importance in attempts to recover any historical signal^{22–26}.

The problems associated with phylogenetic reconstruction come into particularly sharp focus when comparing support for the three-domains and eocyte trees. The first studies to investigate this question generally recovered the three-domains tree, in which eukaryotes emerge as the sister group to Archaea^{27,28}, but the parsimony and distance methods used carried unrealistic assumptions, including constancy (homogeneity) of base compositions across lineages and of evolutionary rates across sites. These assumptions are clearly violated by key phylogenetic markers such as small subunit ribosomal RNA genes, which contain a mixture of fast- and slowly-evolving sites²⁹ and for which GC content varies widely among the three domains¹². Compositional heterogeneity can cause phylogenetic error when not taken into account, because sequences of similar base or amino acid composition may group together in the tree even when they are not closely related^{30–32}. Two pioneering studies used methods to mitigate possible convergence in the universal tree due to shared compositional biases in nucleotide sequences and, interestingly, both recovered an eocyte tree^{33,34}.

Long branch attraction (LBA) is another pervasive artefact in molecular phylogenies, in which sequences with long branches cluster together irrespective of their evolutionary history^{25,35}. LBA is especially problematic for parsimony methods, but it can also affect probabilistic methods if the model ignores among-site rate variation, or is otherwise a poor fit to the data³⁶. Trees for the rRNA and protein-coding genes used to infer relationships between domains often show evidence of long branches and are therefore susceptible to LBA. Some of the early attempts to mitigate the influence of LBA in inter-domain analyses also recovered eocyte trees, although with variable support. Evolutionary parsimony, a method designed to reduce the effect of long branches on the inferred tree, recovered an eocyte topology from rRNA sequences³⁷, although archaeal monophyly was favoured when a related method, compositional statistics, was used to analyse RNA polymerase sequences³⁸. By contrast, analyses of rRNA and RNA polymerase using models that accounted for among-site rate variation supported the eocyte hypothesis over the three-domains tree³⁹. To reconcile these results, Tourasse and Gouy³⁹ suggested that the three-domains tree might be a phylogenetic artefact caused by LBA between the long bacterial and eukaryotic branches, forcing an artefactual clustering of the shorter archaeal branches. In other words, the eocyte tree might be intrinsically more difficult to recover using simple methods, because it requires the clustering of the short branch leading to the eocytes/Crenarchaeota with the long eukaryotic branch.

Single-gene phylogenies often fail to strongly resolve the relationships between the domains^{12,40}, and so a number of studies have analysed concatenations of the core set of proteins conserved on all genomes. As already described, these genes largely function in translation and gene expression, and include many of the essential RNA and protein components of the ribosome. These cellular components have been called the ‘genealogy-defining core’³, the ‘genetic core’⁴¹ of cells or the ‘functional core of genomes’¹⁶, and their common history has been cited^{3,16,41} as the strongest support for the three-domains tree. Testing the evolutionary origins of this small set of genes is therefore critical to the three-domains–eocyte debate. Interestingly, analyses using similar sets of concatenated core genes have yielded different conclusions, for example Katoh *et al.*⁴² obtained an eocyte tree from a set of 39 universal proteins, whereas Ciccarelli *et al.*⁴³ analysed a similar set of proteins and obtained a three-domains tree. One reason for the conflicting results in this case may be the different methods used for making the sequence alignment: the order of alignment had previously been shown to dictate which tree (eocyte or three-domains) was recovered from elongation factor Tu sequences⁴⁴. Ciccarelli *et al.*⁴³ aligned bacterial, archaeal and eukaryotic sequences separately before combining

them into a single alignment. This stepwise procedure was criticized as potentially biasing the results towards a three-domains topology but also, when the individual alignments were combined, to have introduced alignment errors between domains¹⁹. Brown *et al.*⁴⁵ also inferred trees from a concatenation of a subset of 14 universally conserved proteins, but in this study the tree recovered depended on the phylogenetic method used; the three-domains topology was recovered using maximum parsimony, but model-based methods recovered an eocyte topology.

Over the past few years, phylogenetic models implemented in either a maximum likelihood or a Bayesian framework have continued to increase in sophistication by incorporating additional features of the evolutionary process. These include relaxing the assumptions of homogeneous amino acid or base composition across sites⁴⁶ or across branches of the tree³¹. These models seem to fit molecular sequence data much better than simpler models and this may make them less susceptible to LBA and other artefacts of model mis-specification²⁵. Although relatively few analyses of the core gene set have used these models so far, all of them have recovered the eocyte tree, rather than the three-domains tree^{12,22,47–49}.

New archaeal lineages and eukaryotic origins

In addition to improvements in phylogenetic methods, the diversity of molecular sequences from organisms related to the eocytes has also increased considerably, driven by the ease with which sequences from uncultured prokaryotes can now be sampled from the environment using molecular methods^{50,51}. Improved sampling can have positive effects on phylogenetic reconstruction, particularly when it helps to break up long branches⁵². Recently discovered relatives of the eocytes include the Korarchaeota⁵³, the Thaumarchaeota⁵⁴ and the Aigarchaeota⁵⁵; the ‘TACK’ superphylum was subsequently proposed as an informal group to encompass these four taxa⁴⁷. To date, the studies including TACK sequences have supported a version of the eocyte hypothesis extended to recognize this improved sampling, rather than the three-domains tree^{47,48,56}. In this extended sense, the eocyte hypothesis implies that the closest relative of the eukaryotic nuclear lineage is one, or all, of the TACK Archaea. If this tree is correct, then an important place to look for prokaryotic homologues of eukaryotic cellular componentry should be among the TACK phyla. Consistent with this prediction, members of this group encode homologues of a number of key eukaryotic genes (Fig. 2 and Supplementary Table 1), including actin⁵⁷ and tubulin⁵⁸ (the essential components of the eukaryotic cytoskeleton), a ubiquitin protein-modification system⁵⁵ and a number of genes involved in transcription and translation^{47,59}. However, no single characterized TACK genome possesses all of these features^{47,57,58}, implying that gene loss, and potentially horizontal gene transfer (HGT), have contributed to the patterns of gene sharing on contemporary archaeal and eukaryotic genomes^{60,61}.

Which history do universal trees represent?

In their seminal papers, Woese and Fox^{1,4} recognized that the rRNA tree represented only one component, the host for the mitochondrial endosymbiont, in the composite origins of the eukaryotic cell. That composite nature has been confirmed by comparative genomics, which has demonstrated that eukaryotic genomes contain a mixture of genes with different origins^{13,62–66}. Some genes are ancestrally present in all three groups or unique to eukaryotes, but many others appear to have origins through gene transfers from different Bacteria, including the endosymbiotic progenitors of mitochondria and plastids, and relatively few—including the core set of conserved proteins we have been discussing—have affinities with the Archaea. From these data it is clear that no one tree is sufficient to describe the history of all of the genes on modern eukaryotic genomes^{67,68}. However, even though this fact is now widely documented, the three-domains tree is often still called the ‘tree of life’ or ‘universal tree’ in textbooks¹⁵ and reviews^{16,17}.

The sequencing of genomes from across the tree of eukaryotes is beginning to provide a clearer picture of the impact on eukaryotic genomes of HGT from prokaryotes⁶⁵. These data suggest that the acquisition of bacterial genes, at least by microbial eukaryotes, has been an ongoing

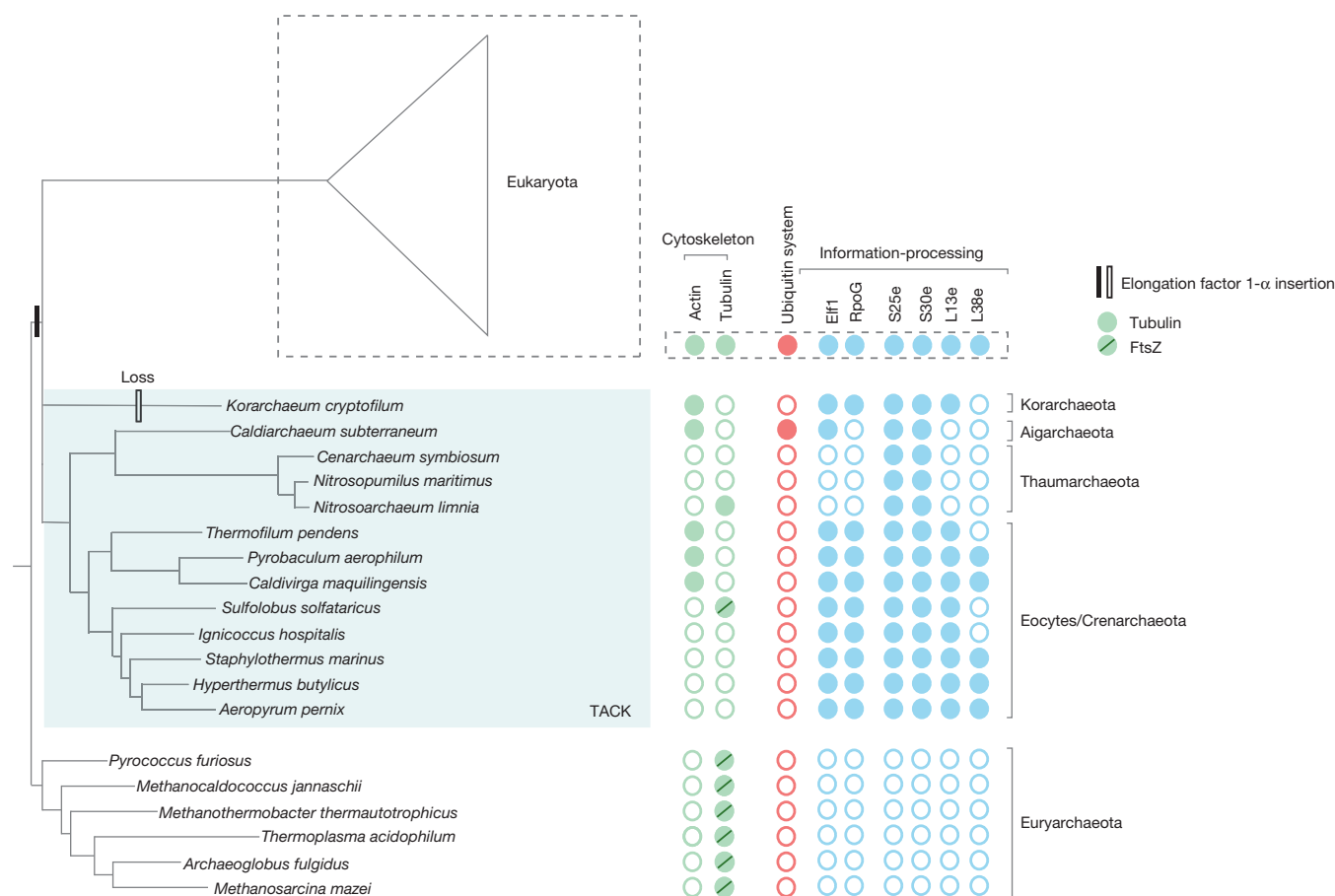


Figure 2 | Archaeal links in the origin of eukaryotes. A schematic tree depicting the relationships between Archaea and the eukaryotic nuclear lineage, consistent with recent analyses of core genes using new methods^{47–49} and rooted using the Bacteria as the outgroup^{5–11}. The phylogenetic position of *Korarchaeum* was not consistently resolved in these different analyses and hence is depicted as part of a polytomy. Genome analyses have detected homologous genes in Archaea and eukaryotes that are consistent with them sharing a common ancestor to the exclusion of Bacteria. Many of these patterns of gene sharing do not distinguish between the rooted three-domains or eocyte trees, as they are expected to occur under both hypotheses. Recently published analyses of the genomes of TACK Archaea, however, have increased the number of homologues shared with eukaryotes and some of these are relevant to ideas about eukaryotic origins and the evolution of their unique features. These include putative orthologues of actin⁵⁷ and tubulin⁵⁸, which in eukaryotes

form the core of the cytoskeleton, as well as components of a ubiquitin protein-modification system in *Caldiarchaeum subterraneum*⁵⁵. Distant homologues of some of these genes have also been detected in Euryarchaeota^{104,105}, but they cluster outside the eukaryote/TACK clade in phylogenetic trees^{57,58,106}. We have followed existing usage⁵⁸ in distinguishing between the FtsZ-like tubulin family members found in some Archaea and the eukaryote-like tubulin homologue found in *Nitrosoarchaeum*. Several eukaryotic genes involved in transcription and translation have prokaryotic homologues or conserved sequence features that have been found so far only among the TACK Archaea. These include four ribosomal proteins⁴⁷, the RNA polymerase subunit RpoG⁵⁹, the elongation factor Elf1 (ref. 107), and a short amino acid insertion¹⁰⁸ in the broadly conserved elongation factor 1- α that has only been found in TACK Archaea and eukaryotes as indicated by the vertical bar. Accession numbers and additional details are provided in Supplementary Tables 1 and 2.

process that extends beyond the initial injection of genes provided by the mitochondrial and plastid endosymbionts. From the perspective of ongoing HGT, the existence of any coherent vertical signal for ancient relationships may seem surprising. However, the impact of HGT on the core genes used to reconstruct the tree of life appears rather limited. Although cases of HGT have been reported^{69,70}, these occur mainly within rather than between domains, and at present there is little evidence that they have generally perturbed inferences of inter-domain relationships^{3,12,41,69}. In addition to genuine cases of HGT, poorly fitting phylogenetic models may also lead to disagreements between gene trees^{25,26}: recent work has shown that improving the fit of phylogenetic models⁴⁸ or integrating the signal from different genes through joint inference of gene and species trees^{71,72} can reduce the level of incongruence and the number of inferred HGT events.

The reasons why core genes involved in transcription, translation, and related processes might be transferred (that is, fixed) less frequently than genes for metabolic pathways are currently understood in terms of their degree of functional integration into cells. Their gene products are often found in large subcellular complexes and therefore tend to have

more interaction partners than genes for metabolic pathways; as a result, horizontal replacement of these genes is more likely to disrupt important cellular interactions and thus to be opposed by negative selection^{66,73,74}. In essence, the universal core might be the largest coherent set of vertically inherited genes that can be tracked across the history of cellular life³, and as such represents a key resource for tracing the emergence of the eukaryotic cellular lineage. Under the rooted three-domains hypothesis¹⁴, that ancestral lineage is as old as the Archaea. By contrast, the eocyte hypothesis predicts that eukaryotes are a relatively young group because their core genes originated from within the Archaea¹⁸.

The origin of eukaryotes in light of other data

In principle, it might be possible to determine the order of events relevant to eukaryotic origins, or at least to exclude some scenarios, using the fossil and biogeochemical record. However, this record is very incomplete and subject to deep and sometimes heated controversy. The first fossil that is indisputably eukaryotic is of a bangiophyte red alga dated to between 1.2 billion and 720 million years ago⁷⁵, but earlier microfossils with a possible eukaryotic origin are found in rocks dated to approximately 1.8 billion

years ago²⁰. These data are consistent with molecular dating analyses that place the last common ancestor of eukaryotes at between 1.9 and 1.7 billion years ago⁷⁶. An earlier origin for eukaryotes had been suggested on the basis of the presence of sterane biomarkers in 2.7-billion-year-old rocks⁷⁷, but these were subsequently shown to be contaminants from younger rocks^{78,79}. An early origin for Archaea has been inferred on the basis of the presence of biological methane, today produced only by methanogenic Euryarchaeota, in rocks that are 3.5 billion years old⁸⁰. Analyses of microfossils and stromatolites—modern versions of which harbour complex bacterial communities⁸¹—in 3.4-billion-year-old rocks suggest the presence of photosynthetic bacteria^{82–84}. Thus, on the data available, Bacteria and Archaea may pre-date eukaryotes in the fossil record by almost 2 billion years.

In light of the uncertainties for dating eukaryotic origins in the geological record, much attention has focused on the historical record revealed by the ultrastructure of the eukaryotic cell and in particular on the timing of the mitochondrial endosymbiosis⁸⁵. When the three primary kingdoms and three-domains tree were originally proposed^{1,14} some contemporary eukaryotes called ‘archezoans’^{85,86} were hypothesized to descend from eukaryotic lineages that never had mitochondria^{85,86}, providing modern-day evidence for the emergence of nucleated cells before the mitochondrial endosymbiosis. The archezoans included the obligate intracellular parasites Microsporidia and a number of parasitic microaerophilic protists including *Entamoeba*, *Giardia* and *Trichomonas*^{85,86}. However, representatives of all of these groups have now been shown to possess a mitochondrial homologue, either a hydrogenosome or mitosome, sharing common ancestry with classical mitochondria^{2,87}. These results imply that the mitochondrion was acquired before the radiation of known eukaryotes; therefore, the observation that the mitochondrion descends from an endosymbiotic member of the alphaproteobacteria^{64,88} provides strong evidence that the origin of eukaryotes postdates the origin of that bacterial group^{2,89}. A relatively late origin of eukaryotes compared to Bacteria is consistent with the best evidence from the geological record and with either the three-domains or eocyte tree rooted on the bacterial stem or within the Bacteria^{5–11}. Moreover, if all eukaryotes have both mitochondria and a nucleus, then we can no longer be sure which structure arose first during evolution: in other words, the host cell that acquired the mitochondrion need not have already possessed a nucleus. Indeed, there are now well-argued hypotheses suggesting that the acquisition of the mitochondrion was the key event that sparked the prokaryote-to-eukaryote transition^{90,91}. In any case, the failure of the Archezoa hypothesis removes a key obstacle to theories that propose a prokaryotic host for the mitochondrial endosymbiont, including hypotheses that are consistent with the eocyte tree².

The origin of eukaryotic cell membranes

The plasma membranes of Bacteria and eukaryotes predominantly contain phospholipids in which fatty acids are covalently bound to *sn*-glycerol-3-phosphate via an ester linkage. By contrast, Archaea—including the few TACK Archaea studied so far—predominantly contain phospholipids with isoprenoid chains linked to *sn*-glycerol-1-phosphate via an ether bond^{92,93}. This pattern is most parsimoniously explained on the rooted three-domains tree by inferring a switch to using mainly glycerol isoprenoid ethers along the archaeal stem, with eukaryotes retaining the ancestral type. This transition may have been driven by a need to maintain membrane function at the high temperatures and acidic conditions of the habitats occupied by early Archaea^{92,94}. A commonly voiced challenge to the eocyte hypothesis—and all archaeal host models for eukaryotic origins—is how to explain the reversion of the archaeal-host membrane to a bacterial-type plasma membrane.

In fact, most of the genes needed for the synthesis of both types of lipid are common to all three groups, suggesting that neither the transition from ester to ether lipids in the common ancestor of Archaea, nor the subsequent reversion along the eukaryotic stem, would require radical genomic change^{95,96}. Archaeal-type ether lipids have been detected in some Bacteria and phospholipids based on *sn*-glycerol-1-phosphate are found

in certain endomembrane components of eukaryotes, suggesting that the distinctions among contemporary membranes may not be as sharp as once thought; there is still much to be discovered about the natural diversity of lipid membranes^{93,95–98}. Moreover, recent experiments have indicated that artificial membranes containing mixtures of bacterial and archaeal lipids are stable⁹⁹, demonstrating the potential for natural mixed-membrane intermediate stages. Given these considerations, the reversion to bacterial-type membranes in eukaryotes might be explained as part of the same process whereby ancestral archaeal pathways were replaced by bacterial equivalents to yield the metabolic similarities observed between Bacteria and contemporary eukaryotes^{62–65,95,96,100}. This transition need not have greatly affected membrane function: in the Haloarchaea, which have obtained a large number of bacterial genes by HGT, transporters derived from Bacteria appear to function normally in the archaeal plasma membrane¹⁰¹.

Conclusions

Ancient phylogenies provide a fascinating window into the distant past, but are difficult to build and interpret – as evidenced by the first thirty years of debate over the tree of life in the era of molecular phylogenetics. Evolutionary biologists now have access to more data and better phylogenetic methods than ever before, although there is still much room for improvement and many uncertainties remain. These caveats apply equally to all attempts to infer ancient relationships, affecting not only the debate over whether the three-domains or eocyte tree best depicts the history of core eukaryotic genes, but also the placement of the universal root^{5,9–12,21} and the relationships among major eukaryotic phyla^{26,102,103}. The pioneering analyses of molecular sequence data led by Carl Woese and his co-workers culminated in the three-domains tree recognizing the Archaea, Bacteria and Eukaryota as the three primary domains of cellular life. Although evidence of widespread HGT means that no single tree can depict the history of all genes on prokaryotic and eukaryotic genomes, the three-domains tree holds a special place in biology. It appears in most textbooks and reviews, where it is often called the ‘universal tree’ and the ‘tree of life’. But support for the iconic three-domains tree has waned with improvements in phylogenetic methods and taxon sampling. Within the limits of methods and data, a version of the eocyte tree is now the best-supported hypothesis for the origin of the subset of genes that mainly function in translation and appear to be most resistant to HGT. The placement of these genes, and by extension the eukaryotic nuclear lineage, within the Archaea is consistent with only two primary lineages and with hypotheses for a symbiogenic origin for eukaryotes involving an archaeon and one or more bacterial partners. The eocyte tree, if correct, suggests that the TACK Archaea, currently a relatively unexplored group, might contain additional clues as to the origin of complex eukaryotic structures. It also rejects the hypothesis that eukaryotes are a primordial cellular lineage, leaving only two candidate primary domains, Archaea and Bacteria, and it identifies a key piece of the puzzle—the host lineage—in the chimaeric origins of the eukaryotic cell.

Received 17 June; accepted 14 October 2013.

1. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
A landmark paper that, together with ref. 4, reported the discovery of the Archaea and discussed its far-reaching implications for early evolution.
2. Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
3. Woese, C. R. On the evolution of cells. *Proc. Natl Acad. Sci. USA* **99**, 8742–8747 (2002).
4. Woese, C. R. & Fox, G. E. The concept of cellular evolution. *J. Mol. Evol.* **10**, 1–6 (1977).
5. Doolittle, W. F. & Brown, J. R. Tempo, mode, the progenote, and the universal root. *Proc. Natl Acad. Sci. USA* **91**, 6721–6728 (1994).
6. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359 (1989).
Together with ref. 7, this paper presented the first evidence for rooting the tree of life on the bacterial stem, but see ref. 5 for a still-relevant discussion of these analyses and other contemporary ideas about early evolution.

7. Gogarten, J. P. *et al.* Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665 (1989).
8. Dagan, T., Roettger, M., Bryant, D. & Martin, W. Genome networks root the tree of life between prokaryotic domains. *Genome Biol. Evol.* **2**, 379–392 (2010).
9. Lake, J. A., Skophammer, R. G., Herbold, C. W. & Servin, J. A. Genome beginnings: rooting the tree of life. *Phil. Trans. R. Soc. B* **364**, 2177–2185 (2009).
10. Skophammer, R. G., Servin, J. A., Herbold, C. W. & Lake, J. A. Evidence for a gram-positive, eubacterial root of the tree of life. *Mol. Biol. Evol.* **24**, 1761–1768 (2007).
11. Cavalier-Smith, T. Rooting the tree of life by transition analyses. *Biol. Direct* **1**, 19 (2006).
12. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaeobacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105**, 20356–20361 (2008).
- The first of a series of recent papers demonstrating that analyses of core genes using new phylogenetic models favour the eocyte tree rather than the three-domains tree.**
13. Doolittle, W. F. & Zhaxybayeva, O. in *The Prokaryotes: Prokaryotic Biology and Symbiotic Associations* (ed. Rosenberg, E.) (Springer, 2013).
- A very clear discussion about the issues facing the integration of phylogenetics and classification given the evidence for extensive lateral gene transfer.**
14. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Woese and colleagues present their arguments for the rooted three-domains tree of life.**
15. Madigan, M. T., Marting, J. M., Stahl, D. A. & Clark, D. P. *Brock Biology of Microorganisms* 13th edn (Benjamin Cummings, 2010).
16. Pace, N. R. Time for a change. *Nature* **441**, 289 (2006).
17. Pace, N. R. Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* **73**, 565–576 (2009).
18. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81**, 3786–3790 (1984).
- This paper presents comparisons of ribosomal structure in Bacteria, Archaea and eukaryotes, providing the initial motivation for the eocyte hypothesis.**
19. Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. & Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nature Rev. Microbiol.* **8**, 743–752 (2010).
20. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. R. Soc. B* **361**, 1023–1038 (2006).
21. Philippe, H. & Forterre, P. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49**, 509–523 (1999).
22. Foster, P. G., Cox, C. J. & Embley, T. M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Phil. Trans. R. Soc. B* **364**, 2197–2207 (2009).
23. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. Mathematical elegance vs biochemical realism: the covarian model of molecular evolution. *J. Mol. Evol.* **53**, 711–723 (2001).
24. Ho, S. Y. & Jermin, L. Tracing the decay of the historical signal in biological sequence data. *Syst. Biol.* **53**, 623–637 (2004).
25. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7** (suppl. 1), S4 (2007).
26. Philippe, H. *et al.* Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
27. Gouy, M. & Li, W. H. Phylogenetic analysis based on rRNA sequences supports the archaeobacterial rather than the eocyte tree. *Nature* **339**, 145–147 (1989).
28. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
29. Olsen, G. J. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 825–837 (1987).
30. Foster, P. G. & Hickey, D. A. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48**, 284–290 (1999).
31. Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).
32. Hirt, R. P. *et al.* Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
33. Lake, J. A. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl Acad. Sci. USA* **91**, 1455–1459 (1994).
34. Yang, Z. & Roberts, D. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* **12**, 451–458 (1995).
- An important early contribution demonstrating that modelling changing nucleotide composition in RNA sequences from different species supported the eocyte tree.**
35. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
36. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nature Rev. Genet.* **13**, 303–314 (2012).
37. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186 (1988).
38. Sidow, A. & Wilson, A. C. Compositional statistics: an improvement of evolutionary parsimony and its application to deep branches in the tree of life. *J. Mol. Evol.* **31**, 51–68 (1990).
39. Tourasse, N. J. & Gouy, M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phylogenet. Evol.* **13**, 159–168 (1999).
40. Yutin, N., Makarova, K. S., Mekhedov, S. L., Wolf, Y. I. & Koonin, E. V. The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.* **25**, 1619–1630 (2008).
41. Harris, J. K., Kelley, S. T., Spiegelman, G. B. & Pace, N. R. The genetic core of the universal ancestor. *Genome Res.* **13**, 407–412 (2003).
42. Katoh, K., Kuma, K. & Miyata, T. Genetic algorithm-based maximum-likelihood analysis for molecular phylogeny. *J. Mol. Evol.* **53**, 477–484 (2001).
43. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).
44. Lake, J. A. The order of sequence alignment can bias the selection of tree topology. *Mol. Biol. Evol.* **8**, 378–385 (1991).
45. Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28**, 281–285 (2001).
46. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
- One of the most notable improvements in phylogenetic modelling in the last decade, providing a Bayesian framework for accommodating across-site compositional heterogeneity—a key feature of molecular sequence data.**
47. Guy, L. & Ettema, T. J. The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).
48. Williams, T. A., Foster, P. G., Nye, T. M., Cox, C. J. & Embley, T. M. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. R. Soc. Lond. B* **279**, 4870–4879 (2012).
49. Lasek-Nesselquist, E. & Gogarten, J. P. The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol. Phylogenet. Evol.* **69**, 17–38 (2013).
50. Pester, M., Schleper, C. & Wagner, M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* **14**, 300–306 (2011).
51. Lloyd, K. G. *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013).
52. Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**, 9–17 (1998).
53. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 8102–8107 (2008).
54. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Rev. Microbiol.* **6**, 245–252 (2008).
55. Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39**, 3204–3223 (2011).
56. Kelly, S., Wickstead, B. & Gull, K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. R. Soc. Lond. B* **278**, 1009–1018 (2011).
57. Ettema, T. J., Lindas, A. C. & Bernander, R. An actin-based cytoskeleton in archaea. *Mol. Microbiol.* **80**, 1052–1061 (2011).
58. Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biol. Direct* **7**, 10 (2012).
59. Koonin, E. V., Makarova, K. S. & Elkins, J. G. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerase are conserved in hyperthermophilic Crenarchaeota and ‘‘Korarchaeota’’. *Biol. Direct* **2**, 38 (2007).
60. Csürös, M. & Miklos, I. Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* **26**, 2087–2095 (2009).
61. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**, 46 (2012).
62. Ribeiro, S. & Golding, G. B. The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15**, 779–788 (1998).
- Together with ref. 63, this paper presented some of the first tree-based evidence that eukaryotes are genomic chimaeras containing some genes that are most similar to those of Bacteria and others to Archaea.**
63. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
64. Esser, C. *et al.* A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
65. Alsmark, C. *et al.* Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol.* **14**, R19 (2013).
66. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaeobacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl Acad. Sci. USA* **107**, 17252–17255 (2010).
67. Dagan, T. & Martin, W. The tree of one percent. *Genome Biol.* **7**, 118 (2006).
68. Doolittle, W. F. & Bapteste, E. Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl Acad. Sci. USA* **104**, 2043–2049 (2007).
69. Williams, D. *et al.* A rooted net of life. *Biol. Direct* **6**, 45 (2011).
70. Creevey, C. J., Doerks, T., Fitzpatrick, D. A., Raes, J. & Bork, P. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLoS ONE* **6**, e22099 (2011).
71. Boussau, B. *et al.* Genome-scale coestimation of species and gene trees. *Genome Res.* **23**, 323–330 (2013).
72. Szöllösi, G. J., Boussau, B., Abby, S. S., Tannier, E. & Daubin, V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc. Natl Acad. Sci. USA* **109**, 17513–17518 (2012).
73. Cohen, O., Gophna, U. & Pupko, T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28**, 1481–1489 (2011).

74. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806 (1999).
 75. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).
 76. Parfrey, L. W., Lahr, D. J., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108**, 13624–13629 (2011).
 77. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
 78. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455**, 1101–1104 (2008).
 79. Fischer, W. W. Biogeochemistry: life before the rise of oxygen. *Nature* **455**, 1051–1052 (2008).
 80. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archean era. *Nature* **440**, 516–519 (2006).
 81. Papineau, D., Walker, J. J., Mojzsis, S. J. & Pace, N. R. Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. *Appl. Environ. Microbiol.* **71**, 4822–4832 (2005).
 82. Allwood, A. C. *et al.* Controls on development and diversity of Early Archean stromatolites. *Proc. Natl Acad. Sci. USA* **106**, 9548–9555 (2009).
 83. Tice, M. M. & Lowe, D. R. Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* **431**, 549–552 (2004).
 84. Schopf, J. W. Fossil evidence of Archean life. *Phil. Trans. R. Soc. B* **361**, 869–885 (2006).
 85. Cavalier-Smith, T. Eukaryotes with no mitochondria. *Nature* **326**, 332–333 (1987).
 86. Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H.E.A.) 1027–1034 (De Gruyter, 1983).
 87. van der Giezen, M., Tovar, J. & Clark, C. G. Mitochondria-derived organelles in protists and fungi. *Int. Rev. Cytol.* **244**, 175–225 (2005).
 88. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
 89. Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. Lond. B* **263**, 1053–1059 (1996).
 90. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
 91. Martin, W. & Koonin, E. V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
 92. Lombard, J., Lopez-Garcia, P. & Moreira, D. The early evolution of lipid membranes and the three domains of life. *Nature Rev. Microbiol.* **10**, 507–515 (2012).
 93. Pitcher, A. *et al.* Core and intact polar glycerol dibiphytanyl glycerol tetraether lipids of ammonia-oxidizing archaea enriched from marine and estuarine sediments. *Appl. Environ. Microbiol.* **77**, 3468–3477 (2011).
 94. van de Vossenberg, J. L., Driessen, A. J. & Konings, W. N. The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2**, 163–170 (1998).
 95. Boucher, Y., Kamekura, M. & Doolittle, W. F. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* **52**, 515–527 (2004).
 96. Lombard, J., Lopez-Garcia, P. & Moreira, D. An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol. Biol. Evol.* **29**, 3261–3265 (2012).
 97. Guldán, H., Matysik, F. M., Bocola, M., Sterner, R. & Babinger, P. Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in bacteria. *Angew. Chem. Int. Edn Engl.* **50**, 8188–8191 (2011).
 98. Tan, H. H., Makino, A., Sudesh, K., Greimel, P. & Kobayashi, T. Spectroscopic evidence for the unusual stereochemical configuration of an endosome-specific lipid. *Angew. Chem. Int. Edn Engl.* **51**, 533–535 (2012).
 99. Shimada, H. & Yamagishi, A. Stability of heterochiral hybrid membrane made of bacterial sn-G3P lipids and archaeal sn-G1P lipids. *Biochemistry* **50**, 4114–4120 (2011).
- Reports the production of stable heterochiral membranes containing a mixture of bacterial- and archaeal-type lipids, demonstrating the feasibility of natural mixed membranes.**
100. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
 101. Nelson-Sathi, S. *et al.* Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl Acad. Sci. USA* **109**, 20537–20542 (2012).
 102. Hampl, V. *et al.* Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic ‘‘supergroups’’. *Proc. Natl Acad. Sci. USA* **106**, 3859–3864 (2009).
 103. Song, S., Liu, L., Edwards, S. V. & Wu, S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl Acad. Sci. USA* **109**, 14942–14947 (2012).
 104. Lindås, A. C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. & Bernander, R. A unique cell division machinery in the Archaea. *Proc. Natl Acad. Sci. USA* **105**, 18942–18946 (2008).
 105. Makarova, K. S., Yutin, N., Bell, S. D. & Koonin, E. V. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nature Rev. Microbiol.* **8**, 731–741 (2010).
 106. Blombach, F. *et al.* Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol. Direct* **4**, 39 (2009).
 107. Daniels, J. P., Kelly, S., Wickstead, B. & Gull, K. Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* **4**, 24 (2009).
 108. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a Marie Curie postdoctoral fellowship to T.A.W. T.M.E. acknowledges support from the European Research Council Advanced Investigator Programme and the Wellcome Trust. We thank J. Archibald for comments on the manuscript.

Author Contributions T.A.W., P.G.F., C.J.C. and T.M.E. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.E. (Martin.Embley@ncl.ac.uk).

Ferromagnetism in suspensions of magnetic platelets in liquid crystal

Alenka Mertelj¹, Darja Lisjak¹, Miha Drofenik^{1,2} & Martin Čopič^{1,3}

More than four decades ago, Brochard and de Gennes proposed that colloidal suspensions of ferromagnetic particles in nematic (directionally ordered) liquid crystals could form macroscopic ferromagnetic phases at room temperature. The experimental realization of these predicted phases has hitherto proved elusive, with such systems showing enhanced paramagnetism but no spontaneous magnetization in the absence of an external magnetic field. Here we show that nanometre-sized ferromagnetic platelets suspended in a nematic liquid crystal can order ferromagnetically on quenching from the isotropic phase. Cooling in the absence of a magnetic field produces a polydomain sample exhibiting the two opposing states of magnetization, oriented parallel to the direction of nematic ordering. Cooling in the presence of a magnetic field yields a monodomain sample; magnetization can be switched by domain wall movement on reversal of the applied magnetic field. The ferromagnetic properties of this dipolar fluid are due to the interplay of the nematic elastic interaction (which depends critically on the shape of the particles) and the magnetic dipolar interaction. This ferromagnetic phase responds to very small magnetic fields and may find use in magneto-optic devices.

A rich variety of complex self-assembled structures of colloidal particles in nematic liquid crystals that are the result of nematic-mediated anisotropic interaction between the particles is of wide-ranging interest for both fundamental science and applications^{1–4}. Nematic liquid crystals are particularly interesting because of their coupling with external electric and magnetic fields. Ferromagnetic ordering in fluids has been experimentally observed in only two systems: liquid helium⁵ and undercooled alloys at temperatures above 1,000 K (ref. 6). But more than 40 years ago, Brochard and de Gennes suggested that a true fluid ferromagnetic phase could appear at room temperature in a colloid of magnetic nanoparticles in nematic liquid crystal⁷. In such a system, orientational order of the nematic liquid crystal would impose orientational order on the anisotropic magnetic particles and, as in ferrofluids in an external magnetic field, magnetic ordering would also appear. Polar nematic ordering was first treated theoretically in ref. 8.

Ferrofluids are stable colloidal dispersions of monodomain magnetic nanoparticles in isotropic liquids, and are well known. Monodomain nanoparticles act as nanomagnets that begin to align in small magnetic fields. The magnetic interaction between the nanoparticles in ferrofluids is tuned so that in the absence of an external magnetic field, the average magnetic interaction of randomly oriented magnetic moments of particles at contact is smaller than thermal energy $k_B T$, so that the particles remain dispersed owing to entropic forces. When an external magnetic field is applied, magnetic moments on average orient along the external magnetic field. Because oriented magnetic dipoles attract head-to-tail, chaining of the particles along the external field occurs, leading to the well known increase of viscosity of a ferrofluid^{9,10}. But whereas a ferrofluid becomes anisotropic only in an external field, the symmetry of the nematic phase of a liquid crystal is uniaxial. Nematic liquid crystals are composed of rod-shaped molecules that in the nematic phase on average orient along a common direction, called the director, that is usually denoted by a unit vector \mathbf{n} with inversion symmetry $\mathbf{n} \equiv -\mathbf{n}$. The orientation direction can be controlled by external electric or magnetic fields or by preparation of confining surfaces, which can impose a defined orientation. Particles that are suspended in a liquid

crystal and have a given orientation of \mathbf{n} at the surface induce a deformation of orientation in the surrounding liquid crystal. This elastic distortion exerts a force on neighbouring particles at a range up to a few micrometres. In addition, defect lines or points can appear around particles. This liquid-crystal-mediated interaction can have either dipolar nature (when the defect is point-like) or quadrupolar nature (in the case of a line defect surrounding the particle) and can be either repulsive or attractive. Polydispersity, different shapes and/or different surfaces of the particles result in complicated structural forces between the particles, and only in simple cases have these been evaluated^{11–13}.

Magnetic nanoparticles in liquid crystal

A suspension of ferromagnetic nanoparticles in the isotropic phase of a liquid crystal behaves as any other ferrofluid. In the nematic phase, however, if the shape of the particles is anisotropic, they adopt a certain orientation with respect to the nematic order. For particles that are on average either prolate or oblate, nematic order has an effect similar to that of an external magnetic field, but with an important difference: because of the equality of the \mathbf{n} and $-\mathbf{n}$ directions, the orientational ordering of the particles induced by the nematic phase is not enough for macroscopic magnetization to appear. For ferromagnetic ordering, a sufficiently strong magnetic interaction between the particles is necessary, but that causes aggregation, so a nematic-mediated repulsive interaction is needed. A single elongated particle, prepared so that the orientation of a nematic director $\mathbf{n}(\mathbf{r})$ (where \mathbf{r} denotes the position vector in the medium) at its surface is parallel to the surface and to the particle's long axis, orients itself with its long axis parallel to \mathbf{n}_0 if placed in a nematic of average director orientation \mathbf{n}_0 . The deformation of the nematic field is small and the resulting nematic interaction is too weak to prevent aggregation. Rod-shaped particles also tend to get next to each other, which leads to antiparallel magnetic dipoles. It has been shown¹⁴ that, in a mixture of elongated ferrite particles in *N*-(4-methoxybenzylidene)-4-butaniline (MBBA), magnetic hysteresis is obtained in a magnetic field of about 1 mT. The sample used in ref. 14, however, was spherical, with a diameter of 6 mm, the coercive

¹J. Stefan Institute, SI-1000 Ljubljana, Slovenia. ²Faculty of Chemistry and Chemical Engineering, University of Maribor, SI-2000 Maribor, Slovenia. ³Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia.

field depended on sample size, and no attempt was made to characterize either the orientation of the nematic or—in particular—the organization of the particles, that is, whether the particles were aggregated. In later experiments^{15–19}, aggregation was avoided by using only very diluted stable suspensions of elongated ferromagnetic particles in nematic liquid crystals, where the magnetic interaction between the particles was negligible and, consequently, in the absence of external magnetic field, macroscopic magnetization was not observed. The suspensions, however, showed paramagnetic behaviour, and in many cases the reorientation of the nematic director was observed at magnetic fields that were lower than the field needed to reorient pure liquid crystal.

If, on the other hand, the elongated particle is prepared so that the orientation of $\mathbf{n}(\mathbf{r})$ at its surface is normal to that surface, then it orients itself with its long axis normal to \mathbf{n}_0 , and the particle's long axis can rotate in the plane perpendicular to \mathbf{n}_0 . The nematic-mediated interaction is weakly attractive and such that the relative orientation between pairs of particles depends on their relative positions, again not stabilizing ferromagnetic order. In the first stable suspension of elongated particles treated for perpendicular surface orientation, the particles were oriented perpendicularly to the nematic director and the observed effects were independent of the direction of the external magnetic field¹⁵. The application of magnetic field along the director caused reversible flocculation of the particles, which was most likely to have been the result of field-induced distortion of nematic orientation that broke the rotational symmetry of the nematic phase and caused monopolar interaction between the particles²⁰.

Ferromagnetic order and magnetic domains

We show now that by using nanometre-sized ferromagnetic platelets with surface treatment that favours a perpendicular orientation of the nematic director at the particle surface, a stable nematic suspension can be produced with macroscopic spontaneous magnetization along the nematic director \mathbf{n}_0 . This is the result of subtle interplay of the nematic-mediated force and the magnetic interaction between the particles. A platelet-shaped particle with planar surfaces treated such that $\mathbf{n}(\mathbf{r})$ at the surface is normal to the surface plane will orient itself with its surface normal parallel to \mathbf{n}_0 . Pairs of such particles interact as quadrupoles (Fig. 1a). The quadrupolar nematic force between the particles gives the strongest attraction when the line joining two particles makes an angle of about 50° with respect to \mathbf{n}_0 . At this angle, the interaction between two magnetic dipoles is small but still such that it favours parallel, that is ferromagnetic, ordering of the dipoles (Fig. 1a).

Preparation of a stable isotropic suspension of magnetic platelets is more demanding in comparison with spherical or elongated particles: this is due to large polydispersity and the fact that platelets form very stable aggregates because of their shape²¹. Recently, stable suspensions of scandium-doped barium hexaferrite single-crystal nanoplatelets in 1-butanol have been prepared with a narrower distribution of platelet size²² (Fig. 1b), and we used these in our experiments. Barium hexaferrite has high magnetocrystalline anisotropy with a preferred direction of the magnetization along the c axis, which is perpendicular to the plane of the platelets. The thickness of the platelets is about 5 nm, and the distribution of the platelet diameter is approximately log-normal, with mean of 70 nm and standard deviation of 38 nm.

A suspension of magnetic platelets with number concentrations in the range 10^{13} – 10^{14} cm^{−3} in liquid crystalline material pentylycyano-biphenyl (5CB) in the isotropic phase was used to fill planar glass cells; the interior surfaces of the cells imposed homogeneous orientation of the director in the plane of the sample. When the cells were slowly cooled, we observed many elongated aggregates forming at the nematic–isotropic phase boundary. This can be avoided by fast quenching into the nematic phase. Depending on platelet concentration, some sparse aggregates still remain (Extended Data Figs 1 and 2). Polarizing microscopy revealed that the nematic director in all samples is well aligned. The samples in the nematic phase are stable, and no additional aggregation occurs after several months and many exposures to external

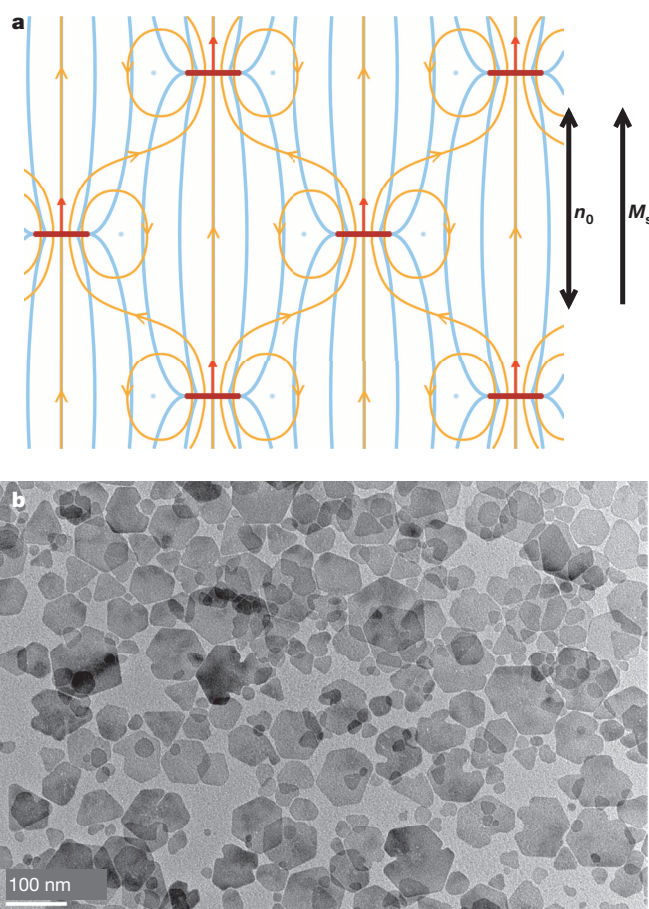


Figure 1 | Magnetic nanoplatelets. **a**, A schematic presentation of the distortion of the director (blue) and magnetic field (orange) around disk-like platelets (short, thick horizontal lines) represented side-on. Blue dots indicate cross-sections of disclination lines, and red arrows the directions of magnetic moments. The average director \mathbf{n}_0 and spontaneous magnetization \mathbf{M}_s parallel to it are also shown. **b**, TEM image of magnetic platelets.

magnetic fields. When a small external magnetic field is applied, the aggregates in slowly cooled cells orient along the field direction while the surrounding nematic liquid crystal remains unchanged. The situation is completely different in quenched cells, where at a field of approximately 5–10 mT, growth of magnetic domains can be observed. Initially the domains are quite small, around 5 μm , then they grow up to a size of a few hundred micrometres (Extended Data Fig. 3). Further observations show that the response of the domains to an external magnetic field depends on the sign of the field, that is, the inversion symmetry of the nematic phase is broken, which is an indication of the spontaneous magnetization. Only two types of domains are observed. When the field is applied along the director, switching is observed only in one type of domain, while the other type of domain remains unchanged. If the direction of the field is reversed, the domains previously unchanged switch and those that switched before remain unchanged (Fig. 2b and c, Extended Data Fig. 1). If the field is applied in the direction perpendicular to the director, both types of domains switch and the domain walls are clearly visible (Fig. 2d and Extended Data Fig. 1). This behaviour shows that the spontaneous magnetization is along the director, and that the two types of domain have magnetization in opposite directions.

Magnetic hysteresis

Monodomain samples (confirmed to be single domain by polarizing microscopy; Extended Data Fig. 2) can be obtained by quenching the suspension from the isotropic to the nematic phase in an external magnetic field that is parallel to the nematic director. The magnetization

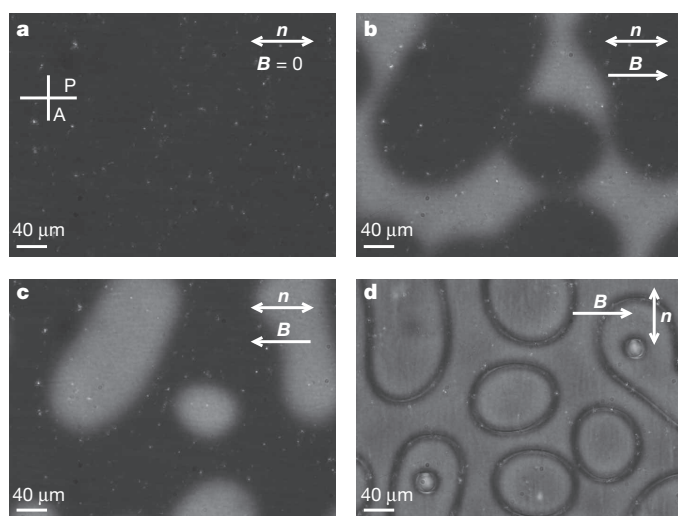


Figure 2 | Switching of ferromagnetic domains in an external magnetic field as seen by polarizing microscopy. The orientation of polarizer P and analyser A for all images is shown in **a**. **a**, Image of an ordered suspension of nanoplatelets in the absence of an external magnetic field ($B = 0$). **b**, **c**, A magnetic field of 3.2 mT is applied along the director (\mathbf{n}) in opposite directions. **d**, A magnetic field of 16 mT is applied perpendicular to the nematic director. The dark lines are domain walls. Platelet concentration in 5CB is 0.16 wt%.

curves of such samples were measured using a vibrating-sample magnetometer (LakeShore 7400 Series VSM). From the magnetization curves we subtracted the diamagnetic contribution measured in a cell filled with pure 5CB, and obtained the magnetization curve belonging to the magnetic platelets (Fig. 3). In the absence of the external field, the magnetic moments of all platelets are oriented in the direction of the field used during sample preparation, and an external magnetic field applied in the same direction does not appreciably change the magnetization of the sample. When the field is applied in the opposite direction, magnetization along the director starts to decrease, reverses and then saturates. The absolute value of saturated magnetization is, within experimental error, the same as the magnetization at zero field, which means that all platelets have rotated by 180° . When the magnitude of the external field is slowly decreased, the magnetization returns to its initial zero-field value. A magnetic hysteresis is observed. At small concentrations the

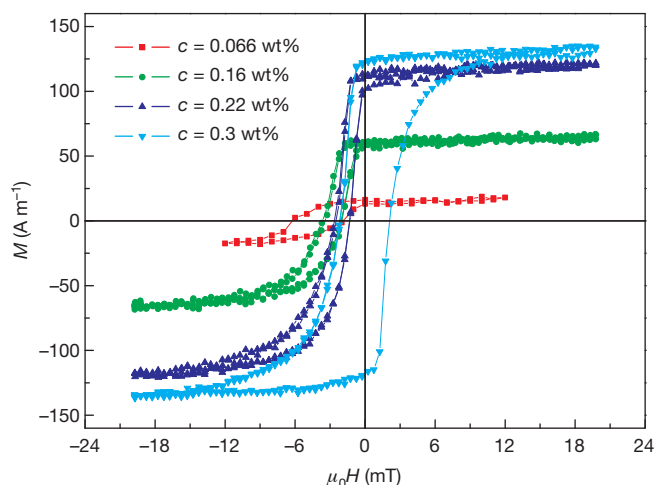


Figure 3 | Magnetization curves of monodomain samples show the switching behaviour of magnetic nanoplatelets in ordered nematic suspension. Magnetization curves are shown for four different platelet concentrations c (key at top left). At lower platelet concentrations, the magnetization in zero field returns to its initial value, whereas in the sample with concentration 0.3 wt%, the magnetization curve is centred around zero.

curves are shifted to negative values of H , whereas for a concentration of 0.3 wt% the curve becomes symmetric.

Observation of switching behaviour under a polarizing microscope (Extended Data Fig. 5) also shows that the nematic director is disturbed by the magnetic field. A field that is initially dark (the darkness showing that the optical axis, which is parallel to \mathbf{n} , is well defined) becomes brighter as the field is reversed, showing that in the field-reversed state \mathbf{n} is distorted and cannot completely reverse owing to the orienting torques at the cell surfaces. At a higher concentration of magnetic platelets in a reversed field of about 10 mT, a system of two white lines crosses the optical field, switching it to the dark state (Fig. 4), that is, we get a complete reversal of the director. Pinning of the director at the cell surfaces causes the observed asymmetry of the hysteresis at lower concentrations. Such optical switching, occurring in a field that is antiparallel with the domain orientation, is unlike the usual Fredericks transition, which is a quadratic field effect in which the director field changes its orientation only when a field larger than a finite critical field is applied perpendicularly to \mathbf{n} . It also proves that coupling with the field is linear, as expected in a ferromagnet.

Shape matters

To explain the mechanism that prevents aggregation and produces spontaneous magnetization, we note the following. The structure of the nematic director field around a single disk-like platelet with strong perpendicular anchoring is of quadrupole symmetry with a defect line—that is, a disclination line (also called a Saturn ring) around the platelet (Fig. 1a). The exact structure of the nematic field depends on the size of the particle and the strength of the anchoring. The disclination line may not be present for very small platelet sizes, but in any case the leading interparticle interaction is quadrupolar. Recent molecular simulations showed that the Saturn ring can be present even around a spherical particle of the size of a few molecular lengths²³. Experiments on colloidal suspensions of gold platelets of similar dimensions also show defects around the rim that are consistent with a Saturn disclination line²⁴. Owing to the anchoring, the nanoplatelets in our system orient so that their plane is perpendicular to the nematic director \mathbf{n}_0 , and, consequently, magnetic moments are parallel to \mathbf{n}_0 . Magnetic interaction is mainly dipolar, favouring particles being vertically in register, while the nematic elastic interaction is quadrupolar, favouring an off-axis

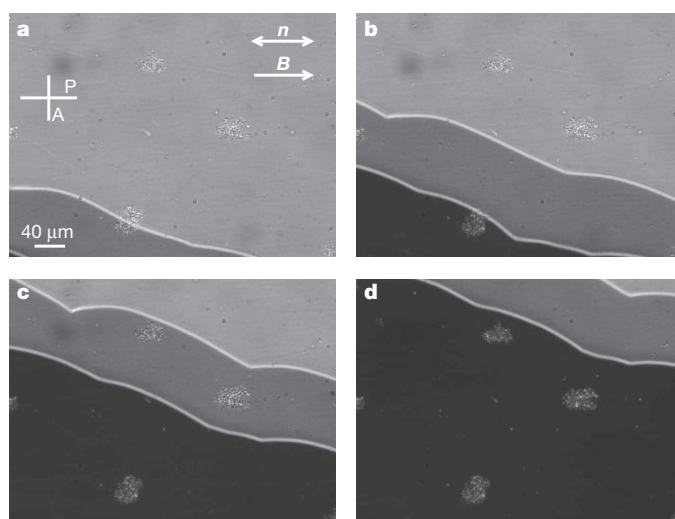


Figure 4 | Time sequence of images showing a complete switching of a monodomain sample. Time increases from **a** to **d**; time between images is 5 s. Images were obtained using polarizing microscopy; P and A in **a** show orientations of polarizer and analyser, respectively, for all panels; also for all panels, a field of 8.4 mT is applied parallel to \mathbf{n} in the opposite direction to the magnetization. Travelling white lines are the surface domain walls, where the director rotates by π . The concentration of the platelets in 5CB is 0.3 wt%.

arrangement (Fig. 1a). The competition between forces reduces the total force to less than $k_B T$, but still induces ferromagnetic ordering by the magnetic interaction. This interplay of magnetic and elastic interactions works for plate-like particles and not for elongated ones. The platelets are also of different diameters, and polydispersity may also play a crucial role in the stability of the suspension, as has been shown in ferrofluids^{25,26}. It has already been shown that molecules with oblate shape, with the dipoles along the short axis, enhance the degree of field-induced polarization in poled polymers²⁷.

Coupling of director and magnetization

Direct magnetic coupling of the nematic liquid crystal with the magnetic field is very weak, because the magnetic anisotropy of 5CB is very small, $\chi_a = 1.6 \times 10^{-6}$, where χ_a is the difference of magnetic susceptibility parallel to and perpendicular to \mathbf{n} . However, the direction of the magnetization is connected with the orientation of the platelets, which is (through surface interaction) coupled to the orientation of the director \mathbf{n} . In the Landau–de Gennes description, the free energy density f of a nematic liquid crystal in an external magnetic field that includes the lowest order of coupling between the director and magnetization can be written as^{28,29}:

$$f = f_{\text{nem}} + \frac{\alpha}{2} M^2 + \frac{\beta}{4} M^4 - \mu_0 \mathbf{M} \cdot \mathbf{H} - \frac{1}{2} \gamma \mu_0 (\mathbf{n} \cdot \mathbf{M})^2 - \frac{1}{2} \chi_a \mu_0 (\mathbf{n} \cdot \mathbf{H})^2 \quad (1)$$

where the first term on the right hand side is the usual free energy density of nematic phase²⁸ f_{nem} , and the following three terms together give the magnetic free energy density: μ_0 is the vacuum permeability, and α and β are the Landau expansion coefficients describing the ferromagnetic transition. The last two terms describe the coupling of the nematic order with magnetization and external magnetic field, respectively. Whereas the direct coupling of \mathbf{n} with the external field \mathbf{H} is weak, its coupling with the magnetization \mathbf{M} described by the constant γ can be much larger.

The behaviour of M with H , and the response to H under the polarizing microscope, can be explained by noting that the nematic director is locked in a given orientation by the orienting surface. So only the interior of the sample switches above a critical field while in a surface layer the director twists back to the initial orientation. The threshold field for the beginning of the reversal of M_s is $B_c = \pi^2 \gamma \mu_0 K M_s / (K \pi^2 + \gamma \mu_0 M_s^2 d^2)$ (Methods, equation (5)), where d is the thickness of the sample, M_s the spontaneous magnetization and K the elastic constant associated with the distortion of \mathbf{n} ; in our case, B_c is around 1 mT. This gives an estimate for γ of ~ 100 . The orientation of \mathbf{M}_s at the surface can switch either by breaking of surface anchoring (homogeneously) or by movement of surface domain π -walls (that can nucleate heterogeneously at sample edges or imperfections) in which the director turns by π . The critical field B_s needed to switch the surface orientation in the absence of nucleated π -walls can be estimated by noting that at B_s the magnetic coherence length $\sqrt{K/\gamma \mu_0 M_s B_s}$ must be of the order of the extrapolation length K/W , where W is the surface anchoring energy. This gives B_s around 100 mT for typical values of K and W , considerably more than what is observed, so that π -wall movement dominates (Fig. 4, Extended Data Figs 4 and 5). This also explains the observed hysteresis loops. For small concentrations, γM_s is too small to drive the movement of π -walls so surfaces mostly do not switch and the magnetization curves are shifted to negative H . For higher concentrations, γM_s is large enough that the surface can switch and the magnetization curve is symmetric. With π -wall movement a polydomain sample can be transformed to a monodomain one (Extended Data Fig. 4).

The system can be probed further by dynamic light scattering on orientational fluctuations in monodomain samples in external field. The well known turbidity of the nematic phase is caused by orientational fluctuations of the nematic director, which are the fundamental hydrodynamic excitations of director field \mathbf{n} . The relaxation rates of these excitations depend on the viscoelastic properties of the nematic liquid crystal and on external fields. Magnetization adds another degree

of freedom coupled to \mathbf{n} , so that two types of fluctuations exist—one is more nematic-like and the other is more magnetic-like. Both are in principle observable by light scattering, but the nematic-like fluctuation is more pronounced. Whereas in the usual nematic phase the dependence of the relaxation rates on the external fields is quadratic, in the case of the presence of spontaneous magnetization M_s a linear term dominates, which is again an indication that the inversion symmetry of the nematic phase is broken. From the free energy density (equation (1)) we can calculate the relaxation rate of fluctuations in the magnetic field along the director (Methods, equation (4)). Experimental results are shown in Fig. 5. For small H , the relaxation rate depends linearly on the external field. At these magnitudes of magnetic field, the relaxation rate of pure 5CB does not change measurably. From values of spontaneous magnetization obtained from Fig. 3 and from the known K of 5CB, we can again estimate the coupling constant $\gamma \approx 50$ –150, in accordance with the value obtained from magnetization curves.

Conclusion and outlook

We have demonstrated the existence of a ferromagnetic nematic colloidal suspension that was theoretically predicted long ago but not until now experimentally obtained. The system has the standard properties of a ferromagnet: monodomain samples can be prepared by cooling in an external field, and the system shows hysteretic behaviour, magnetization reversal in a flipped external field, and domain walls and domain-wall motion. It is an intriguing liquid that possesses two order parameters \mathbf{n} and \mathbf{M} and can therefore be considered a new class of multiferroic material. Ferromagnetic ordering is made possible by the shape of the magnetic nanoparticles that have the form of thin platelets. The quadrupolar nature of the nematic elasticity-mediated interaction prevents particle aggregation and causes the magnetic dipolar interaction to favour ferromagnetic ordering. The new phase switches at very small magnetic fields and may lead to new magneto-optic devices. Magnetization of the particles is a new degree of freedom that is coupled to the nematic order, which could lead to new effects—for example, oscillating director relaxation and field-dependent viscoelastic properties³⁰. Incorporating the particles in other liquid crystal phases with chiral or smectic order would open up a new field of research, that of magnetic phenomena in complex fluids.

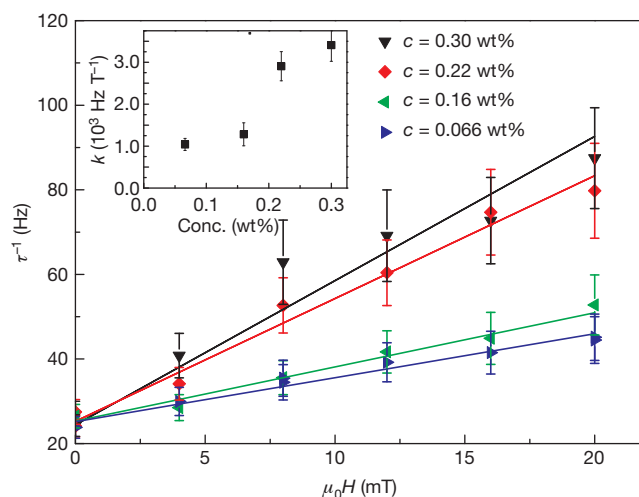


Figure 5 | Dependence of the relaxation rate of director orientational fluctuations on external magnetic field in samples with different concentration of magnetic nanoplatelets. Main panel, data for four different concentrations of platelets are shown (see key): τ^{-1} , relaxation rate. Inset, plot of k versus concentration of platelets (here k is the slope of τ^{-1} versus $\mu_0 H$). The given concentrations include all platelets, both well dispersed and aggregated, so the actual concentration of dispersed platelets is smaller than shown. Error bars, 1σ uncertainty of the least-squares fits.

METHODS SUMMARY

Barium hexaferrite nanoplatelets with the nominal composition $\text{BaFe}_{11.5}\text{Sc}_{0.5}\text{O}_{19}$ were hydrothermally synthesized at 240 °C (ref. 22). The nanoplatelets were dispersed in 1-butanol using dodecylbenzenesulphonic acid as a surfactant²¹. The thickness of the synthesized magnetic platelets is about 5 nm; the distribution of the platelets' diameter is approximately log-normal, with a mean of 70 nm and a standard deviation of 38 nm. Saturated mass magnetization of the platelets measured in the dried suspension was $32 \text{ A m}^2 \text{ kg}^{-1}$, so the magnetic moment of an average platelet is $3 \times 10^{-18} \text{ A m}^2$.

A suspension of 0.6 wt% magnetic platelets in 1-butanol was mixed with liquid crystalline material pentylcyanobiphenyl (5CB, Nematel) in the isotropic phase in different volume ratios. The mixture was kept in the isotropic phase at 50 °C for 24 h to evaporate 1-butanol. The final concentration of platelets in 5CB ranged from 0.06 wt% to 0.6 wt%. Suspensions, still in the isotropic phase, were used to fill planar glass cells, with surfaces imposing parallel orientation of the director (Instec, Inc.). The thickness of the cells was 20.4 μm . The cells were then quenched to give nematic phase. Quenching is necessary because in the samples that were slowly cooled into nematic phase, aggregation of platelets at the nematic–isotropic phase boundary appeared. In samples with a high concentration of platelets many aggregates were visible, however, the surrounding liquid crystal showed strong response to the external field (Extended Data Figs 1 and 2). The number and size of the aggregates decrease with decreasing concentration. At concentrations lower than about 0.06 wt%, the response of the sample to an external magnetic field is weak and inhomogeneous. The nematic–isotropic phase transition temperature of the suspensions is within experimental error (0.1 K) the same as for pure 5CB. Slow heating of the suspensions to give the isotropic phase again causes aggregation of the platelets at the nematic–isotropic boundary.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June; accepted 8 November 2013.

- Poulin, P., Stark, H., Lubensky, T. C. & Weitz, D. A. Novel colloidal interactions in anisotropic fluids. *Science* **275**, 1770–1773 (1997).
- Mušević, I., Škarabot, M., Tkalec, U., Ravnik, M. & Žumer, S. Two-dimensional nematic colloidal crystals self-assembled by topological defects. *Science* **313**, 954–958 (2006).
- Lapointe, C. P., Mason, T. G. & Smalyukh, I. I. Shape-controlled colloidal interactions in nematic liquid crystals. *Science* **326**, 1083–1086 (2009).
- Lavrentovich, O. D. Liquid crystals, photonic crystals, metamaterials, and transformation optics. *Proc. Natl Acad. Sci. USA* **108**, 5143–5144 (2011).
- Paulson, D. N. & Wheatley, J. C. Evidence for electronic ferromagnetism in superfluid $^3\text{He-A}$. *Phys. Rev. Lett.* **40**, 557–561 (1978).
- Albrecht, T. et al. First observation of ferromagnetism and ferromagnetic domains in a liquid metal. *Appl. Phys. A* **65**, 215–220 (1997).
- Brochard, F. & de Gennes, P. G. Theory of magnetic suspensions in liquid crystals. *J. Phys.* **31**, 691–708 (1970).
- Born, M. Über anisotrope Flüssigkeiten. Versuch einer Theorie der flüssigen Kristalle und des elektrischen Kerr-Effekts in Flüssigkeiten. *Sitz. Kön. Preuss. Akad. Wiss.* **30**, 614–650 (1916).
- Ilg, P. & Odenbach, S. in *Colloidal Magnetic Fluids: Basics, Development and Application of Ferrofluids* (ed. Odenbach, S.) 249–326 (Springer, 2009).
- Rosensweig, R. E. *Ferrohydrodynamics* 237–269 (Dover, 1997).
- Senyuk, B. et al. Shape-dependent oriented trapping and scaffolding of plasmonic nanoparticles by topological defects for self-assembly of colloidal dimers in liquid crystals. *Nano Lett.* **12**, 955–963 (2012).
- Senyuk, B. & Smalyukh, I. I. Elastic interactions between colloidal microspheres and elongated convex and concave nanoprisms in nematic liquid crystals. *Soft Matter* **8**, 8729–8734 (2012).
- Eskandari, Z., Silvestre, N. M., Tasinkevych, M. & da Gama, M. M. T. Interactions of distinct quadrupolar nematic colloids. *Soft Matter* **8**, 10100–10106 (2012).
- Rault, J., Cladis, P. E. & Burger, J. P. Ferronematics. *Phys. Lett. A* **32**, 199–200 (1970).
- Chen, S.-H. & Amer, N. M. Observation of macroscopic collective behavior and new texture in magnetically doped liquid crystals. *Phys. Rev. Lett.* **51**, 2298–2301 (1983).
- Kopčanský, P. et al. Structural changes in the 6CHBT liquid crystal doped with spherical, rodlike, and chainlike magnetic particles. *Phys. Rev. E* **78**, 011702 (2008).
- Buluy, O. et al. Magnetic sensitivity of a dispersion of aggregated ferromagnetic carbon nanotubes in liquid crystals. *Soft Matter* **7**, 644–649 (2011).
- Podoliak, N. et al. Macroscopic optical effects in low concentration ferronematics. *Soft Matter* **7**, 4742–4749 (2011).
- Podoliak, N. et al. Magnetite nanorod thermotropic liquid crystal colloids: synthesis, optics and theory. *J. Colloid Interf. Sci.* **386**, 158–166 (2012).
- Lev, B. I., Chernyshuk, S. B., Tomchuk, P. M. & Yokoyama, H. Symmetry breaking and interaction of colloidal particles in nematic liquid crystals. *Phys. Rev. E* **65**, 021709 (2002).
- Ovtar, S., Lisjak, D. & Drogenik, M. Barium hexaferrite suspensions for electrophoretic deposition. *J. Colloid Interface Sci.* **337**, 456–463 (2009).
- Lisjak, D. & Drogenik, M. Chemical substitution—an alternative strategy for controlling the particle size of barium ferrite. *Cryst. Growth Des.* **12**, 5174–5179 (2012).
- Moreno-Razo, J. A. et al. Effects of anchoring strength on the diffusivity of nanoparticles in model liquid-crystalline fluids. *Soft Matter* **7**, 6828–6835 (2011).
- Evans, J. S., Beier, C. N. & Smalyukh, I. I. Alignment of high-aspect ratio colloidal gold nanoplatelets in nematic liquid crystals. *J. Appl. Phys.* **110**, 033535 (2011).
- Mahle, S., Ilg, P. & Liu, M. Hydrodynamic theory of polydisperse chain-forming ferrofluids. *Phys. Rev. E* **77**, 016305 (2008).
- López-López, M. T., Zubarev, A. Y. & Bossis, G. Repulsive force between two attractive dipoles, mediated by nanoparticles inside a ferrofluid. *Soft Matter* **6**, 4346–4349 (2010).
- Dalton, L. R., Harper, A. W. & Robinson, B. H. The role of London forces in defining noncentrosymmetric order of high dipole moment–high hyperpolarizability chromophores in electrically poled polymeric thin films. *Proc. Natl Acad. Sci. USA* **94**, 4842–4847 (1997).
- de Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals* 41–162 (Clarendon, 1995).
- Pleiner, H., Jarkova, E., Muler, H. W. & Brand, H. R. Landau description of ferrofluid to ferronematic phase transition. *Magnetohydrodynamics* **37**, 254–260 (2001).
- Jarkova, E., Pleiner, H., Müller, H.-W. & Brand, H. R. Macroscopic dynamics of ferronematics. *J. Chem. Phys.* **118**, 2422–2430 (2003).

Acknowledgements This work was supported by the Slovenian Research Agency (A.M. and M.Č., grant no. P1-0192; D.L. and M.D., grant no. P2-0089-4). We thank the CENN Nanocenter for use of the LakeShore 7400 Series vibrating-sample magnetometer.

Author Contributions A.M. designed the study and performed the experiments; A.M. and M.Č. interpreted results and wrote the paper; and D.L. and M.D. designed and synthesized nanoplatelets, and prepared the suspension of nanoplatelets in isotropic solvent.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (alenka.mertelj@ijs.si).

Ultrafast endocytosis at mouse hippocampal synapses

Shigeki Watanabe¹, Benjamin R. Rost^{2†*}, Marcial Camacho-Pérez^{2*}, M. Wayne Davis¹, Berit Söhl-Kielczynski², Christian Rosenmund² & Erik M. Jorgensen¹

To sustain neurotransmission, synaptic vesicles and their associated proteins must be recycled locally at synapses. Synaptic vesicles are thought to be regenerated approximately 20 s after fusion by the assembly of clathrin scaffolds or in approximately 1 s by the reversal of fusion pores via ‘kiss-and-run’ endocytosis. Here we use optogenetics to stimulate cultured hippocampal neurons with a single stimulus, rapidly freeze them after fixed intervals and examine the ultrastructure using electron microscopy—‘flash-and-freeze’ electron microscopy. Docked vesicles fuse and collapse into the membrane within 30 ms of the stimulus. Compensatory endocytosis occurs within 50 to 100 ms at sites flanking the active zone. Invagination is blocked by inhibition of actin polymerization, and scission is blocked by inhibiting dynamin. Because intact synaptic vesicles are not recovered, this form of recycling is not compatible with kiss-and-run endocytosis; moreover, it is 200-fold faster than clathrin-mediated endocytosis. It is likely that ‘ultrafast endocytosis’ is specialized to restore the surface area of the membrane rapidly.

Recycling of synaptic vesicle membrane and proteins is required to keep membrane surface area constant and ensure efficient neurotransmission during sustained synaptic activity. Classical ultrastructural analysis of frog neuromuscular junctions led to two models of endocytosis. A slow endocytic pathway that takes place distant from active zones via clathrin scaffolds ~20 s after exocytosis was proposed¹. A fast mechanism, now termed kiss-and-run, that retrieves fusing vesicles by reversing their neck was then suggested^{2,3}—a process that takes place within 1 s^{4–6}. Since then, many studies have focused on understanding the molecular mechanisms and the kinetics of endocytosis to distinguish these two models. However, conflicting evidence has accumulated over the past 40 years.

The studies on the molecular mechanisms suggest that proteins associated with clathrin play crucial roles in synaptic vesicle endocytosis^{7,8}. Purified clathrin and its adaptor proteins are sufficient to reconstitute vesicles from brain-derived liposomes⁹. When cleavage by dynamin is disrupted, clathrin-coated pits and coated vesicles accumulate at the plasma membrane of synapses^{10,11}. Likewise, reductions of clathrin-interacting adaptor proteins^{12–15}, membrane-curvature proteins^{16–18} or scaffolding proteins^{19–21} perturb endocytosis at the synapse. On the other hand, the recruitment of clathrin triskelia is known to be very slow—on a time scale of seconds²², making it difficult to reconcile how vesicles can be regenerated when firing rates exceed 100 Hz²³. Moreover, functional vesicles are still regenerated in the absence of clathrin or its adaptor proteins^{13,24}, suggesting that another pathway may be operating at synapses.

The conclusions from studies on kinetics of endocytosis are often contradictory, but in some cases suggest that both kiss-and-run and clathrin-mediated endocytosis are operational at the synapse. Fluid-phase uptake of fluorescent dyes^{5,25,26} or measurements of open times using quantum dots^{4,27} show two kinetic components: fast (1–2 s) and slow (~20 s), although some authors have reported only a single kinetic component²⁸. Dye release of vesicles during exocytosis indicates that some dye is retained in the vesicle and suggests the presence of a

transient fusion pore that opens during kiss-and-run^{4,5,26}. Capacitance measurements from the calyx of Held²⁹, retinal bipolar cells⁶, and hippocampal mossy fibre boutons³⁰ suggest that both fast and slow mechanisms are probably at work. On the other hand, the measurement of protein trafficking using the pH-sensitive fluorescent protein pHluorin suggests that endocytosis of vesicle proteins in mammalian central synapses occurs with a single time constant of 15 s^{31,32}, similar to the time course revealed by ultrastructural analysis of the clathrin-mediated pathway¹.

The classic ultrastructural studies that gave rise to these models have certain caveats. For the cold-glutaraldehyde fixations, minutes- to hours-long stimulations were applied to dissected frog neuromuscular junctions^{1,2}. For the freeze-slammer experiments, 4-aminopyridine was applied to block potassium channels and prolong release^{3,33}. To observe synaptic ultrastructure following a single physiological stimulation, we developed a device that couples optogenetics and rapid high-pressure freezing³⁴. Using this ‘flash-and-freeze’ approach, we found that endocytosis occurs within 50 ms after stimulation at the edges of active zones in the nematode *Caenorhabditis elegans*³⁴. Here we applied a brief single light stimulus to mouse hippocampal neurons expressing channelrhodopsin and froze them at various time points ranging from 15 ms to 10 s after stimulation. Morphometry was conducted blind for 200 synapses from each time point. We found that endocytosis takes place between 50 and 100 ms after stimulation at the edge of active zones (Extended Data Fig. 1). These results indicate that ultrafast endocytosis mediated by actin and dynamin may be the first step in synaptic vesicle recycling in mouse hippocampal neurons.

Docked vesicles fuse in the active zone

To capture membrane dynamics during neurotransmission, we coupled optogenetics with high-pressure freezing of cultured mouse hippocampal neurons. Primary cultures were infected with lentivirus expressing channelrhodopsin-2 (E123T/T159C; hereafter ‘ChetaTC’)^{35,36}. On-cell electrophysiological recordings performed at 34 °C revealed that a single 10 ms pulse of blue light elicits at least one action potential with a mean

¹Department of Biology and Howard Hughes Medical Institute, University of Utah, Salt Lake City 84112, Utah, USA. ²Neuroscience Research Centre, Charité Universitätsmedizin, 10117 Berlin, Germany.

†Present address: German Center for Neurodegenerative Diseases (DZNE), 10117 Berlin, Germany.

*These authors contributed equally to this work.

delay of 4.8 ± 0.4 ms after light onset in 88% of the infected neurons (Fig. 1a and Extended Data Fig. 2). Some cells (20%, 7/34) fired a second action potential within 30 ms of light-on as illustrated in the sample trace (Fig. 1a). No action potential was observed in 12% of the cells (4/34; Extended Data Fig. 2b). Application of tetrodotoxin blocks light-evoked neurotransmitter release (Fig. 1b, upper trace), demonstrating that vesicle fusion is triggered by action potentials rather than by a global membrane depolarization by channelrhodopsin. Postsynaptic currents were observed with a delay of $\sim 6.9 \pm 0.3$ ms after light-on (Fig. 1b), and were blocked by NBQX and bicuculline (Fig. 1b, middle trace). In paired recordings of isolated ChetaTC positive and negative neurons, alternating light or electrical stimulation in the ChetaTC-positive cell elicits identical postsynaptic currents, demonstrating that in our conditions channelrhodopsin is acting solely via action potential generation, rather than supplementing calcium in the synaptic bouton (Extended Data Fig. 3).

To capture synaptic structures during neurotransmission, we applied a single 10-ms pulse of blue light, froze the cells using a high-pressure freezer at defined time points ranging from 15 ms up to 10 s after light onset, and prepared the samples for electron microscopy (Extended Data Fig. 2d, see Methods). To enhance release probability, the external solution contained 4 mM calcium. Omega figures of fusing vesicles were only rarely observed in unstimulated synapses (1% profiles with no stimulation, Fig. 1f), whereas they were found in 20% of synaptic profiles 15 ms after light onset (~ 10 ms after the initiation of an action potential). Thus, these structures probably represent fusions occurring during

peak transmission or phasic release (Fig. 1b). Fusions were observed within the active zone (Fig. 1c–e), and the diameters of these pits were similar to those of synaptic vesicles, and thus each pit represents the fusion of a single synaptic vesicle (Fig. 1g). Based on the wide opening at their neck (Fig. 1c–e), these pits are likely to be collapsing into the membrane, and the omega structures perdure for 20 ms ($\tau = 20.3$ ms, see Methods). Occasionally, multiple pits were observed in single sections (Fig. 1e), indicating that multi-vesicular release can occur in response to a single action potential or tightly spaced action potentials (20% of stimulated cells, Fig. 1a and Extended Data Fig. 2b). In either case, the presence of multiple fusing vesicles indicates that the fusion of a second vesicle is not blocked by the preceding fusion event.

We defined docked vesicles as those in direct contact with the membrane, in addition we scored vesicles with tethers that were within 30 nm of the plasma membrane (Fig. 2a). The number of docked vesicles was reduced from 1.7 to 1.0 per profile after the stimulus, but the tethered pool was largely unaffected (Fig. 2b–d). The docked pool recovered with a time constant of 3.8 s (Fig. 2b), which is similar to the time constant (4.3 s) for the recovery of the readily releasable pool previously measured in mouse hippocampal neurons by electrophysiology³⁷. These data indicate that docked vesicles are likely to be the morphological correlates of the physiologically defined readily-releasable pool.

Endocytosis peaks at 100 ms

Following exocytosis, endocytic structures appeared adjacent to the active zone (Fig. 3 and Fig. 4; Extended Data Fig. 4 for more example electron micrographs). Shallow pits (<40 nm), deep pits (>40 nm), and fully internalized vesicles (>50 nm) were all observed within 50 ms after stimulation and were declining by 300 ms (Fig. 3a–d, i). The number of shallow and deep pits increases again at 1 s, indicating a possible secondary wave of endocytosis. The diameters of these endocytic pits and large vesicles are similar (86 ± 2.4 nm and 82 ± 1.2 nm, respectively), suggesting a precursor–product relationship between these structures (Fig. 3h). The surface area of large vesicles is comparable to the surface area of about four synaptic vesicles (see membrane calculations in Methods). Electron-dense filaments are occasionally visible

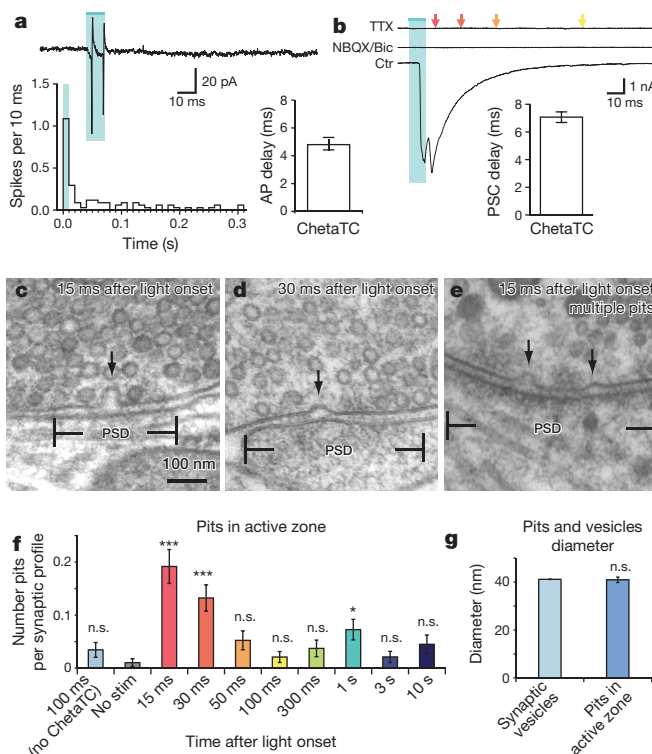


Figure 1 | Channelrhodopsin activation induces action-potential-driven vesicle fusion. **a**, Cell-attached voltage clamp recordings of light-evoked action potentials. The histogram shows the average number of action potentials (AP) at different time points after the light application. Each column is binned by 10 ms. **b**, Sample trace of light-evoked postsynaptic currents. Arrows indicate freezing times at 15, 30, 50, and 100 ms. **c**, **d**, Representative micrographs of fusing vesicles at 15 ms (**c**) and 30 ms (**d**) after the light onset. **e**, A micrograph showing two exocytic intermediates in the active zone. **f**, Average number of pits in the active zone at different time points after stimulation. **g**, Diameter of synaptic vesicles (41.1 ± 0.1 nm; $n = 382$) and pits in active zones (40.9 ± 1.2 nm; $n = 62$; $P = 0.7$). *** $P < 0.0001$; * $P < 0.006$. n.s., not significant. Error bars, s.e.m.

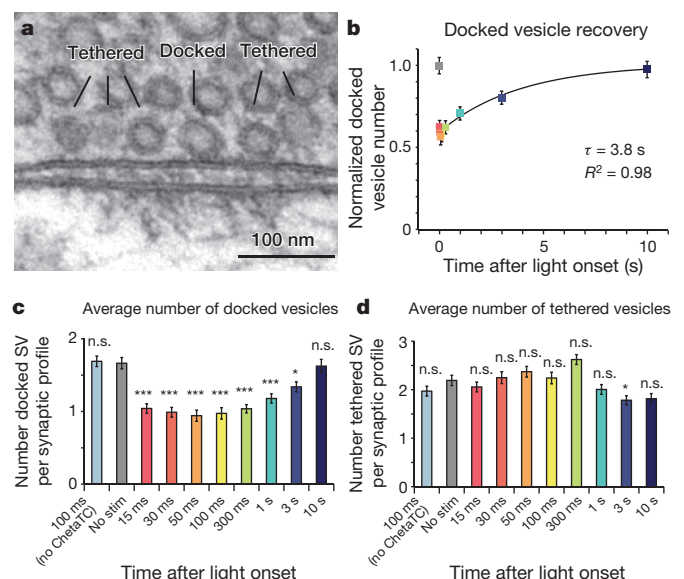


Figure 2 | Docked vesicles are the morphological correlates of the readily-releasable pool. **a**, A sample micrograph showing docked and tethered vesicles. **b**, Recovery of docked vesicles. The number of docked vesicles is normalized to the non-stimulated control (0 ms). The τ for recovery was 3.8 s. **c**, **d**, Average number of synaptic vesicles (SV) docked (**c**) or tethered (**d**) in the active zone at time points after the stimulus. For detailed numbers and statistical analysis, see Supplementary Table 1. *** $P < 0.0001$; * $P < 0.006$. n.s., not significant. Error bars, s.e.m.

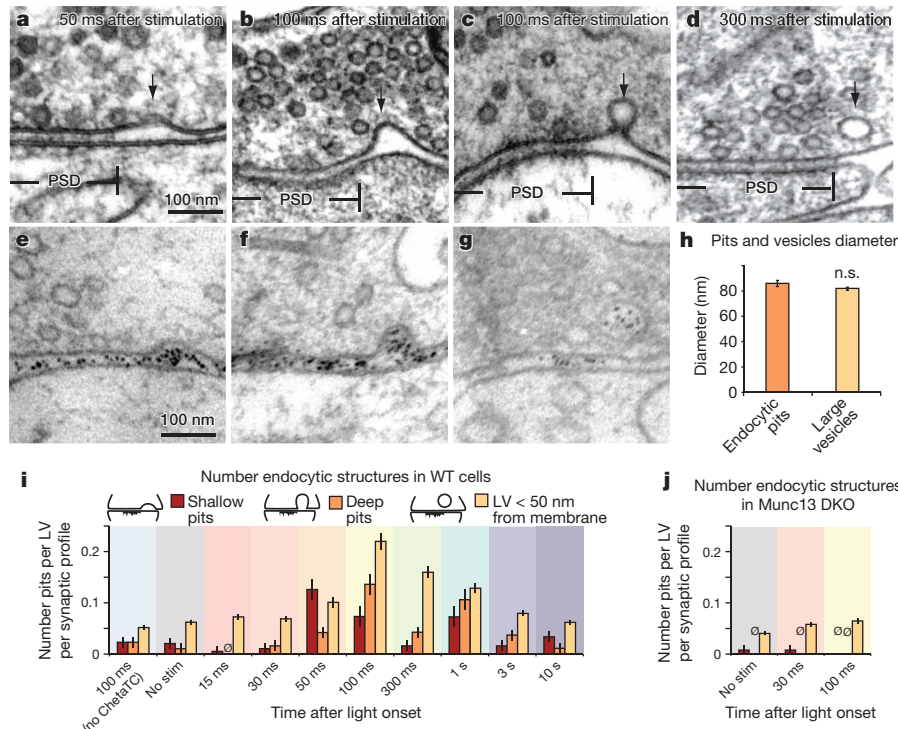


Figure 3 | Large invaginations immediately next to the active zone are endocytic intermediates. **a–d**, Representative micrographs showing invaginations and large vesicles in the periactive zone at 50 ms (**a**), 100 ms (**b**, **c**) and 300 ms (**d**) after stimulation. **e–g**, Sample micrographs showing ferritin uptake into these large endocytic structures at 100 ms after stimulation. **h**, Diameter of large vesicles (82.0 ± 1.2 nm; $n = 572$) and endocytic invaginations (86.0 ± 2.4 nm; $n = 51$; $P = 0.33$). **i**, Number of

shallow pits (<40 nm), deep pits (>40 nm) and large vesicles (LV; <50 nm from plasma membrane) in the periactive zone at different time points after stimulation. **j**, Number of shallow pits, deep pits and large vesicles in the periactive zone at different time points after stimulation in Munc13 DKO. For detailed numbers and statistical analysis, see Supplementary Table 1. n.s., not significant. Error bars, s.e.m.

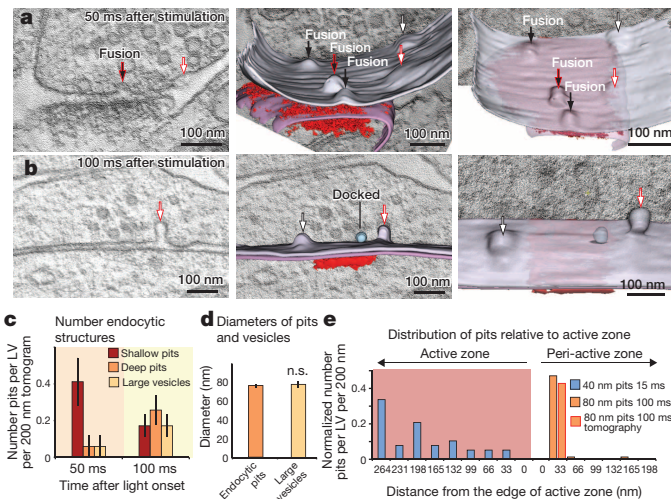


Figure 4 | Endocytosis is localized to sites flanking the active zone. **a**, **b**, Tomograms of synapses frozen 50 ms (**a**) or 100 ms (**b**) after stimulation. Left, a virtual 2 nm section from the tomograms. Middle, orthogonal views of the reconstructed volumes. Right, tilted views of the reconstructed volumes. The postsynaptic density is coloured red to delineate the extent of the active zone. Black and white arrows indicate fusing vesicles and endocytic pits, respectively. Arrows outlined in red indicate the plane of the image shown in the left panels. **c**, Number of shallow pits, deep pits and large vesicles in the periactive zone at different time points after stimulation. **d**, Diameter of large vesicles (78.0 ± 3.3 nm; $n = 6$) and endocytic invaginations (76.3 ± 1.9 nm; $n = 15$; $P = 0.56$). **e**, Distribution of pits relative to the edge of active zone. For detailed numbers and statistical analysis, see Supplementary Table 1. n.s., not significant. Error bars, s.e.m.

around these endocytic structures (Extended Data Fig. 4j, k), but they differ from classical coats (Extended Data Fig. 4m–p).

To confirm the spatial relationship of exocytic and endocytic sites, we performed electron tomography on 200-nm thick sections at 50 ms and 100 ms after stimulation (Fig. 4). Fusing vesicles, defined by their diameter, were rare and were not observed in the 100-ms tomograms (Fig. 4a). Endocytic invaginations were outside the active zone and were often observed flanking the active zone (Fig. 4b, e). The number of shallow invaginations declines at 100 ms and is associated with an increase in the number of deep pits and large vesicles (Fig. 4c) similar to the 40-nm sections (Fig. 3i). The diameters of pits (76.3 ± 1.9 nm) and large vesicles (78.0 ± 3.3 nm) are similar, again suggesting a precursor–product relationship (Fig. 4d).

One potential interpretation of these data is that synaptic vesicles first undergo homotypic compound fusion to generate a large vesicle^{38,39}, and then fuse to the plasma membrane. To exclude such a possibility we used ferritin as a fluid phase marker to label endocytic structures. Cationized ferritin binds to negative charges on the plasma membrane and becomes internalized by endocytosis. Ferritin-positive synaptic vesicles were only rarely observed in the absence of stimulation (Extended Data Fig. 5a; control no stimulation, see Methods). Cells were stimulated with a single light pulse and frozen 100 ms later. After stimulation, ferritin was found in invaginations (22% of the synapses) and large vesicles (16% of the synapses; Fig. 3e–g and Extended Data Fig. 5c–e for more example micrographs). The frequencies of endocytic structures were similar to matched controls with no ferritin, thus ferritin itself did not affect ultrafast endocytosis (Extended Data Fig. 5f, g). The presence of ferritin in large vesicles but not synaptic vesicles (Extended Data Fig. 5b) indicates that these large structures are due to endocytosis rather than compound fusion, and further indicates that synaptic vesicles are not being regenerated directly from the membrane on this time scale.

Ultrafast endocytosis scales with exocytosis

Some data suggest that calcium influx may be able to stimulate endocytosis without vesicle fusion⁴⁰. To test if calcium influx alone can stimulate ultrafast endocytosis, we repeated the experiment in cultured neurons from Munc13-1 Munc13-2 double mutants (Munc13-1/2 DKO), which should still experience calcium influx but lack synaptic vesicle exocytosis at hippocampal synapses^{41,42}. We applied a single light pulse to Munc13-1/2 DKO hippocampal neurons expressing ChetaTC and froze the sample at 0 ms, 30 ms and 100 ms. The number of docked vesicles was significantly reduced as previously observed^{41,42}, and no exocytic omega structures were observed. Endocytic structures were completely absent in Munc13-1/2 DKO cells (Fig. 3j). Thus, ultrafast endocytosis requires the addition of membrane to the surface via exocytosis and is not solely driven by calcium.

Nevertheless, increasing calcium is known to accelerate endocytosis during synaptic activity⁴³. In our experiments we used 4 mM calcium to increase release probability; it is possible that ultrafast endocytosis depends on elevated intracellular calcium. We therefore repeated the experiments using 2 mM calcium in the external solution (Extended Data Fig. 6). At least one action potential during the light pulse was observed in 73% of cells (11/15; Extended Data Fig. 6a, b). Endocytic intermediates were captured at the lateral edge of active zone 100 ms after stimulation (Extended Data Fig. 6c–e). The number of exocytic events and endocytic events was lower than the number observed with 4 mM calcium (Extended Data Fig. 6f, g). However, the kinetics of endocytosis were not altered. These results indicate that ultrafast endocytosis occurs under physiological conditions, but scales with the amount of membrane added by exocytosis.

To determine if ultrafast endocytosis can fully compensate for membrane added during exocytosis, we compared the surface area added by exocytosis to that removed by endocytosis (see Methods). The number of fused vesicles was determined by comparing docked vesicles at the 0 ms and 15 ms time points. We calculated that 3,700 nm² of membrane was added to the surface per each profile. The amount of membrane endocytosed at 100 ms was 3,150 nm². In short, we observe that 0.7 synaptic vesicles fuse per profile and 0.6 vesicle equivalents are recovered per profile. These data indicate that a majority of the membrane added by full-collapse fusion after a single stimulus is retrieved by ultrafast endocytosis.

Actin and dynamin mediate endocytosis

Some forms of endocytosis, either clathrin-dependent or -independent, are mediated by actin⁴⁴. To investigate the role of actin polymerization in ultrafast endocytosis, we applied latrunculin-A (Fig. 5), which disrupts polymerization of actin⁴⁵. Application of 0.1% dimethylsulphoxide (DMSO) or 10 μ M latrunculin-A for 1 min did not affect evoked neurotransmission (Fig. 5d and Extended Data Fig. 7a, b). Hippocampal neurons treated with either latrunculin-A or DMSO for 30 s were frozen 100 ms after light stimulation (Fig. 5a, b and Extended Data Fig. 7c–h for more representative micrographs). The DMSO-treated neurons showed normal endocytosis (Fig. 5b, e). No endocytic structures were observed in the latrunculin-A treated cells (Fig. 5a, e), indicating that polymerized actin is required for ultrafast endocytosis.

Cleavage of budded vesicles from the plasma membrane is associated with the GTPase dynamin^{10,11}. Dynasore is a small molecule that interferes with the GTPase activity of dynamin and blocks the cleavage of membrane necks⁴⁶, but may also affect the actin cytoskeleton⁴⁷. We applied 80 μ M dynasore for 30 s before freezing. Unstimulated neurons treated with dynasore did not accumulate pits, indicating that spontaneous activity does not contribute substantially to trapped endocytic intermediates (Fig. 5f; 0.1 pits or large vesicles per profile in both treated and control). In cultures stimulated by a single pulse of light and frozen 100 ms later, dynasore treatment did not increase the total number of endocytic structures, indicating that dynasore did not increase exocytosis and thereby indirectly increase endocytic structures (Fig. 5f; 0.4 endocytic structures per profile in treated and control). However, the

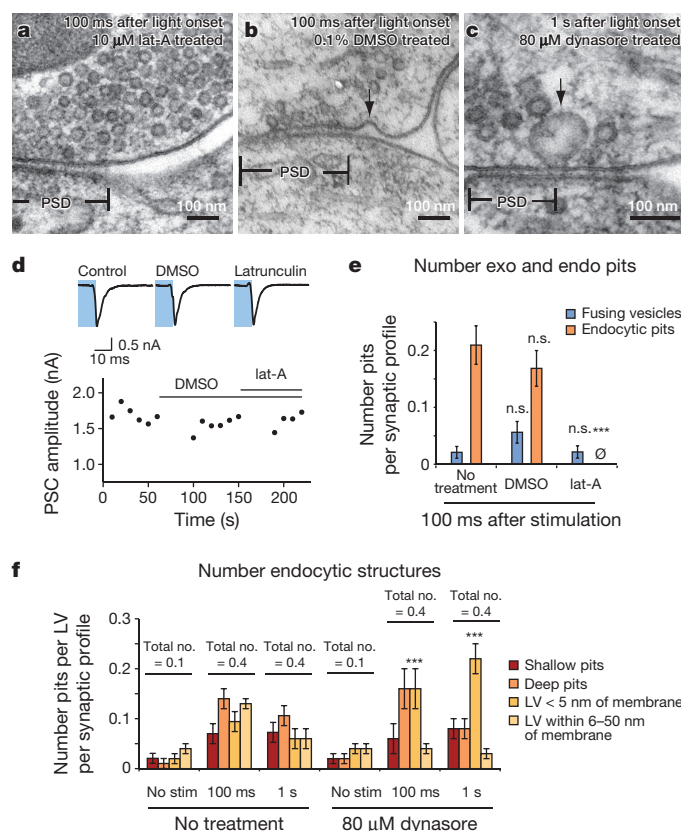


Figure 5 | Ultrafast endocytosis is mediated by actin and dynamin.

a, b, Representative micrographs showing 10 μ M latrunculin-A (**a**) and 0.1% DMSO (**b**) 100 ms after light stimulation. **c**, A representative micrograph showing 80 μ M dynasore-treated cells 1 s after light stimulation. **d**, Traces of average postsynaptic currents in the absence of drugs, during the DMSO application, and during the latrunculin-A application. **e**, Average number of exocytic pits (blue) and endocytic pits (orange) in the latrunculin-A- or DMSO-treated cells. **f**, Number of shallow pits, deep pits and large vesicles associated with the membrane (<5 nm) and large vesicles associated with the active zone (6–50 nm of the membrane). Controls are from the samples in Fig. 3. For detailed numbers and statistical analysis, see Supplementary Table 1. *** $P < 0.0001$. n.s., not significant. Error bars, s.e.m.

fraction of structures remaining as large vesicles trapped close to the membrane increased (Fig. 5f; control: 0.09 large vesicles per profile, treated: 0.16 large vesicles). By 1 s after stimulation, the endocytic structures had progressed to large vesicles, but remained associated with the membrane in dynasore-treated samples (control: 0.06 large vesicles, treated: 0.22 large vesicles; Fig. 5c, f and Extended Data Fig. 8a–c for more micrographs). Necks were only visible in 18% of the trapped large vesicles (10/57), possibly due to the thickness of the sections (40 nm). Electron tomography revealed that the large vesicles were still attached to the membrane (Extended Data Fig. 8d–f), indicating that these large vesicles are trapped on the surface. Together, these results indicate that endocytosis in hippocampal neurons occurs within 100 ms lateral to the active zone, and is mediated by actin polymerization and dynamin.

Discussion

It is widely agreed that after high-frequency stimulation, membranes and proteins are retrieved by clathrin^{7,48}. The best estimates for the speed of clathrin-mediated endocytosis at the synapse come from vesicle proteins tagged with pHluorin. Synaptic vesicle proteins freshly added to the surface after a single action potential are recycled with a time constant of about 15–20 s^{31,32}. In addition, a fraction of vesicle proteins already reside on the plasma membrane and are only endocytosed after stimulation⁴⁹. This readily retrievable pool of proteins is also internalized with a time constant of about 20 s after stimulation⁴⁹. Endocytosis

of synaptic vesicle proteins requires clathrin and clathrin adaptors and is probably specialized to bind and recover synaptic vesicle proteins lost to the plasma membrane³². These rates are slow compared to ultrafast endocytosis, which is 200 times faster, but the assembly kinetics of clathrin are known to be slow²². It is too premature to eliminate a role for clathrin in ultrafast endocytosis; the absence of distinctive clathrin coats does not exclude a role for clathrin in ultrafast endocytosis.

There is a large body of evidence for a fast pathway attributed to kiss-and-run endocytosis^{5,25,26}. Ultrafast endocytosis is not kiss-and-run because of three characteristics: first, the vesicle fusions we observe have broad openings, suggesting they are collapsing into the membrane. Second, ultrafast endocytosis takes place at the lateral edge of the active zone, rather than in the middle of the active zone where fusion takes place. Third, the surface area of the invaginations is equivalent to four synaptic vesicles rather than one vesicle. It is unlikely that these large vesicles represent the recovery of intact synaptic vesicles.

Nevertheless, kiss-and-run events should be captured in our preparation, because they are thought to be more abundant during low-frequency stimulation², and the pore is estimated to last approximately 1–2 s⁵. Thus kiss-and-run vesicles should be visible in the preparation at all time points between 15 ms and 1 s. However, the pore is estimated to be only 1–2 nm in diameter^{2,5}. Although such a pore has been observed previously by electron microscopy^{2,3}, such small structures would be difficult to observe routinely in 40-nm thick sections. We cannot exclude the possibility of kiss-and-run figures in our micrographs. However, it is unlikely that kiss-and-run constitutes a significant fraction of endocytic events in our preparation: presumably kiss-and-run vesicles would become undocked after stimulation, and our calculations of membrane surface area suggest that ultrafast endocytosis can account for most docked vesicles lost to fusion after a single stimulus.

Ultrafast endocytosis is likely to be an evolutionarily conserved process. At the frog neuromuscular junction, depressions adjacent to the active zone were observed in freeze-fracture studies 1 s after stimulation³³; these depressions were interpreted as a potentially fast form of endocytosis. Capacitance measures in retinal bipolar cells indicate that a fast form of endocytosis ($\tau = 1$ s) is clathrin-independent⁶. Ultrastructural studies at *C. elegans* neuromuscular junctions indicate that endocytosis can be as fast as 50 ms following a single stimulus and also occurs at the lateral edges of the active zone³⁴. The data presented here suggest that ultrafast endocytosis almost instantaneously compensates for membrane added to the plasma membrane. The rapid communication between fusion in the active zone and the endocytic sites at the lateral edges is probably mediated by a relaxation in membrane tension. The intriguing unknown is how membrane tension could be coupled to the endocytic machinery.

METHODS SUMMARY

Mouse hippocampal neurons were cultured on astrocyte feeder layers in mass cultures. Cells were infected using lentivirus expressing channelrhodopsin-2 (E123T/T159C). Experiments were performed after 13–15 days in culture. Action potentials were recorded in the cell-attached voltage clamp configuration after a single 10-ms light stimulus in 34 °C HEPES-buffered solution containing 4 mM Ca²⁺, 1 mM Mg²⁺, 3 μ M NBQX and 30 μ M bicuculline.

For measuring synaptic delay and effects of latrunculin-A, we infected 20% of cells with mKate and 80% of cells with ChetaTC, co-cultured the infected cells, and patched mKate-positive cells that lacked photocurrents. Whole-cell patch clamp recordings were conducted at 34 °C in the absence of NBQX or bicuculline.

For electron microscopy, neurons were grown on 6-mm sapphire disks. The neurons were transferred into the same extracellular solutions ~30 s before freezing. A single 10-ms light pulse was applied at defined time points before high-pressure freezing. Following freeze-substitution, en bloc staining with uranyl acetate and plastic embedding, serial 40-nm sections were collected. The samples were blinded before imaging and analysis. Approximately 200 synaptic profiles were randomly chosen from each time point and analysed blind.

For electron tomography, 200-nm thick sections were collected. Tilt series of $\pm 65^\circ$ were acquired, and tomograms generated using IMOD. Membranes were segmented using Amira or TrakEM2.

For ferritin uptake experiments, cells were immersed in solution containing 0.25 mg ml⁻¹ ferritin for 5 min in reduced external calcium solution (1 mM) to suppress spontaneous activity. For latrunculin-A and dynasore experiments, we applied a solution containing 10 μ M latrunculin or 80 μ M dynasore or 0.1% DMSO 30 s before stimulation and freezing.

For a complete description of statistical tests, see Supplementary Table 1. For detailed description of methods and membrane calculations, see Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June; accepted 1 November 2013.

Published online 4 December 2013.

- Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
- Ceccarelli, B., Hurlbut, W. P. & Mauro, A. Depletion of vesicles from frog neuromuscular junctions by prolonged tetanic stimulation. *J. Cell Biol.* **54**, 30–38 (1972).
- Torri-Tarelli, F., Grohovaz, F., Fesce, R. & Ceccarelli, B. Temporal coincidence between synaptic vesicle fusion and quantal secretion of acetylcholine. *J. Cell Biol.* **101**, 1386–1399 (1985).
- Zhang, Q., Cao, Y.-Q. & Tsien, R. W. Quantum dots provide an optical signal specific to full collapse fusion of synaptic vesicles. *Proc. Natl Acad. Sci. USA* **104**, 17843–17848 (2007).
- Richards, D. A., Bai, J. & Chapman, E. R. Two modes of exocytosis at hippocampal synapses revealed by rate of FM1-43 efflux from individual vesicles. *J. Cell Biol.* **168**, 929–939 (2005).
- Von Gersdorff, H. & Matthews, G. Dynamics of synaptic vesicle fusion and membrane retrieval in synaptic terminals. *Nature* **367**, 735–739 (1994).
- Dittman, J. & Ryan, T. A. Molecular circuitry of endocytosis at nerve terminals. *Annu. Rev. Cell Dev. Biol.* **25**, 133–160 (2009).
- Maycox, P. R., Link, E., Reetz, A., Morris, S. A. & Jahn, R. Clathrin-coated vesicles in nervous tissue are involved primarily in synaptic vesicle recycling. *J. Cell Biol.* **118**, 1379–1388 (1992).
- Takei, K. *et al.* Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes. *Cell* **94**, 131–141 (1998).
- Koenig, J. H. & Ikeda, K. Disappearance and reformation of synaptic vesicle membrane upon transmitter release observed under reversible blockage of membrane retrieval. *J. Neurosci.* **9**, 3844–3860 (1989).
- Shupliakov, O. *et al.* Synaptic vesicle endocytosis impaired by disruption of dynamin-SH3 domain interactions. *Science* **276**, 259–263 (1997).
- Gu, M. *et al.* AP2 hemicomplexes contribute independently to synaptic vesicle endocytosis. *eLife* **2**, e00190 (2013).
- Kim, S. H. & Ryan, T. A. Synaptic vesicle recycling at CNS synapses without AP-2. *J. Neurosci.* **29**, 3865–3874 (2009).
- Nonet, M. L. *et al.* UNC-11, a *Caenorhabditis elegans* AP180 homologue, regulates the size and protein composition of synaptic vesicles. *Mol. Biol. Cell* **10**, 2343–2360 (1999).
- Diril, M. K., Wienisch, M., Jung, N., Klingauf, J. & Haucke, V. Stonin 2 is an AP-2-dependent endocytic sorting adaptor for synaptotagmin internalization and recycling. *Dev. Cell* **10**, 233–244 (2006).
- Schmidt, A. *et al.* Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* **401**, 133–141 (1999).
- Verstreken, P. *et al.* Endophilin mutations block clathrin-mediated endocytosis but not neurotransmitter release. *Cell* **109**, 101–112 (2002).
- Anggono, V. *et al.* Syndapin I is the phosphorylation-regulated dynamin I partner in synaptic vesicle endocytosis. *Nature Neurosci.* **9**, 752–760 (2006).
- Jakobsson, J. *et al.* Role of epsin 1 in synaptic vesicle endocytosis. *Proc. Natl Acad. Sci. USA* **105**, 6445–6450 (2008).
- Koh, T.-W., Verstreken, P. & Bellen, H. J. Dap160/intersectin acts as a stabilizing scaffold required for synaptic development and vesicle endocytosis. *Neuron* **43**, 193–205 (2004).
- Marie, B. *et al.* Dap160/intersectin scaffolds the periaction zone to achieve high-fidelity endocytosis and normal synaptic growth. *Neuron* **43**, 207–219 (2004).
- Cocucci, E., Aguet, F., Boulant, S. & Kirchhausen, T. The first five seconds in the life of a clathrin-coated pit. *Cell* **150**, 495–507 (2012).
- Maeno-Hikichi, Y., Polo-Parada, L., Kastanenka, K. V. & Landmesser, L. T. Frequency-dependent modes of synaptic vesicle endocytosis and exocytosis at adult mouse neuromuscular junctions. *J. Neurosci.* **31**, 1093–1105 (2011).
- Sato, K. *et al.* Differential requirements for clathrin in receptor-mediated endocytosis and maintenance of synaptic vesicle pools. *Proc. Natl Acad. Sci. USA* **106**, 1139–1144 (2009).
- Teng, H., Cole, J. C., Roberts, R. L. & Wilkinson, R. S. Endocytic active zones: hot spots for endocytosis in vertebrate neuromuscular terminals. *J. Neurosci.* **19**, 4855–4866 (1999).
- Harata, N. C., Choi, S., Pyle, J. L., Aravanis, A. M. & Tsien, R. W. Frequency-dependent kinetics and prevalence of kiss-and-run and reuse at hippocampal synapses studied with novel quenching methods. *Neuron* **49**, 243–256 (2006).
- Park, H., Li, Y. & Tsien, R. W. Influence of synaptic vesicle position on release probability and exocytotic fusion mode. *Science* **335**, 1362–1366 (2012).
- Ryan, T. A., Smith, S. J. & Reuter, H. The timing of synaptic vesicle endocytosis. *Proc. Natl Acad. Sci. USA* **93**, 5567–5571 (1996).

29. Renden, R. & von Gersdorff, H. Synaptic vesicle endocytosis at a CNS nerve terminal: faster kinetics at physiological temperatures and increased endocytotic capacity during maturation. *J. Neurophysiol.* **98**, 3349–3359 (2007).
30. Hallermann, S., Pawlu, C., Jonas, P. & Heckmann, M. A large pool of releasable vesicles in a cortical glutamatergic synapse. *Proc. Natl Acad. Sci. USA* **100**, 8975–8980 (2003).
31. Balaji, J. & Ryan, T. A. Single-vesicle imaging reveals that synaptic vesicle exocytosis and endocytosis are coupled by a single stochastic mode. *Proc. Natl Acad. Sci. USA* **104**, 20576–20581 (2007).
32. Granseth, B., Odermatt, B., Royle, S. J. & Lagnado, L. Clathrin-mediated endocytosis is the dominant mechanism of vesicle retrieval at hippocampal synapses. *Neuron* **51**, 773–786 (2006).
33. Miller, T. M. & Heuser, J. E. Endocytosis of synaptic vesicle membrane at the frog neuromuscular junction. *J. Cell Biol.* **98**, 685–698 (1984).
34. Watanabe, S. *et al.* Ultrafast endocytosis at *Caenorhabditis elegans* neuromuscular junctions. *eLife* **2**, e00723 (2013).
35. Gunaydin, L. A. *et al.* Ultrafast optogenetic control. *Nature Neurosci.* **13**, 387–392 (2010).
36. Berndt, A. *et al.* High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels. *Proc. Natl Acad. Sci. USA* **108**, 7595–7600 (2011).
37. Pyott, S. J. & Rosenmund, C. The effects of temperature on vesicular supply and release in autaptic cultures of rat and mouse hippocampal neurons. *J. Physiol.* **539**, 523–535 (2002).
38. Matthews, G. & Sterling, P. Evidence that vesicles undergo compound fusion on the synaptic ribbon. *J. Neurosci.* **28**, 5403–5411 (2008).
39. He, L. *et al.* Compound vesicle fusion increases quantal size and potentiates synaptic transmission. *Nature* **459**, 93–97 (2009).
40. Neale, E. A., Bowers, L. M., Jia, M., Bateman, K. E. & Williamson, L. C. Botulinum neurotoxin A blocks synaptic vesicle exocytosis but not endocytosis at the nerve terminal. *J. Cell Biol.* **147**, 1249–1260 (1999).
41. Varoqueaux, F. *et al.* Total arrest of spontaneous and evoked synaptic transmission but normal synaptogenesis in the absence of Munc13-mediated vesicle priming. *Proc. Natl Acad. Sci. USA* **99**, 9037–9042 (2002).
42. Siksou, L. *et al.* A common molecular basis for membrane docking and functional priming of synaptic vesicles. *Eur. J. Neurosci.* **30**, 49–56 (2009).
43. Armbruster, M., Messa, M., Ferguson, S. M., De Camilli, P. & Ryan, T. A. Dynamin phosphorylation controls optimization of endocytosis for brief action potential bursts. *eLife* **2**, e00845 (2013).
44. Mooren, O. L., Galletta, B. J. & Cooper, J. A. Roles for actin assembly in endocytosis. *Annu. Rev. Biochem.* **81**, 661–686 (2012).
45. Spector, I., Shochet, N. R., Kashman, Y. & Groweiss, A. Latrunculin: novel marine toxins that disrupt microfilament organization in cultured cells. *Science* **219**, 493–495 (1983).
46. Macia, E. *et al.* Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* **10**, 839–850 (2006).
47. Yamada, H. *et al.* Dynasore, a dynamin inhibitor, suppresses lamellipodia formation and cancer cell invasion by destabilizing actin filaments. *Biochem. Biophys. Res. Commun.* **390**, 1142–1148 (2009).
48. McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
49. Wienisch, M. & Klingauf, J. Vesicular proteins exocytosed and subsequently retrieved by compensatory endocytosis are nonidentical. *Nature Neurosci.* **9**, 1019–1027 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank D. Lorenz and A. Muenster-Wandowski for providing access to electron microscopes. We would like to thank P. Hegemann and F. Schneider for providing the ChetaTC construct, A. Felies for cell cultures, B. Brokowski for generating lentivirus, E. Hujber for image processing and freezing calculations, C. Ebeling for calculating the time constant for synaptic vesicle collapse, and J. Iwasa for drawing the model figure. We would like to thank C. Tomova and Leica Microsystems for providing us with technical details of the controller of the high-pressure freezer for precise temporal control of light stimulation. We thank EMBO for providing the travel funds (S.W.). The research was funded by the US National Institutes of Health (NS034307; E.M.J.), European Research Council grant (249939 SYNGLUT; C.R.), and German Research Council grants (EXC 257, SFB 665, SFB958; C.R.). E.M.J. is an Investigator of the Howard Hughes Medical Institute and is an Alexander von Humboldt Scholar.

Author Contributions S.W., C.R., and E.M.J. conceived and designed experiments. M.W.D. designed and programmed the light device. S.W. and B.S.-K. performed the freezing experiments. B.R.R. designed lentivirus constructs. B.R.R. and M.C.-P. performed electrophysiology. S.W. performed electron microscopy imaging and analysis. S.W., B.R.R., M.C.-P., C.R. and E.M.J. wrote the manuscript. C.R. and E.M.J. provided funding; experiments were performed at the Charité Universitätsmedizin, Berlin, Germany.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.J. (jorgensen@biology.utah.edu) and C.R. (christian.rosenmund@charite.de).

Targeting *Plasmodium* PI(4)K to eliminate malaria

Case W. McNamara^{1*}, Marcus C. S. Lee^{2*}, Chek Shik Lim³, Siau Hoi Lim³, Jason Roland¹, Advait Nagle¹, Oliver Simon³, Bryan K. S. Yeung³, Arnab K. Chatterjee¹, Susan L. McCormack¹, Micah J. Manary⁴, Anne-Marie Zeeman⁵, Koen J. Decherer⁶, T. R. Santha Kumar², Philipp P. Henrich², Kerstin Gagarin¹, Maureen Ibanez¹, Nobutaka Kato¹, Kelli L. Kuhen¹, Christoph Fischli⁷, Matthias Rottmann^{7,8}, David M. Plouffe¹, Badry Bursulaya¹, Stephan Meister⁴, Lucia Rameh⁹, Joerg Trappe¹⁰, Dorothea Haasen¹⁰, Martijn Timmerman⁶, Robert W. Sauerwein^{6,11}, Rossarin Suwanarusk¹², Bruce Russell^{12,13}, Laurent Renia¹², Francois Nosten^{14,15}, David C. Tully¹, Clemens H. M. Kocken⁵, Richard J. Glynn¹, Christophe Bodenreider³, David A. Fidock^{2,16}, Thierry T. Diagana³ & Elizabeth A. Winzeler^{1,4}

Achieving the goal of malaria elimination will depend on targeting *Plasmodium* pathways essential across all life stages. Here we identify a lipid kinase, phosphatidylinositol-4-OH kinase (PI(4)K), as the target of imidazopyrazines, a new antimalarial compound class that inhibits the intracellular development of multiple *Plasmodium* species at each stage of infection in the vertebrate host. Imidazopyrazines demonstrate potent preventive, therapeutic, and transmission-blocking activity in rodent malaria models, are active against blood-stage field isolates of the major human pathogens *P. falciparum* and *P. vivax*, and inhibit liver-stage hypnozoites in the simian parasite *P. cynomolgi*. We show that imidazopyrazines exert their effect through inhibitory interaction with the ATP-binding pocket of PI(4)K, altering the intracellular distribution of phosphatidylinositol-4-phosphate. Collectively, our data define PI(4)K as a key *Plasmodium* vulnerability, opening up new avenues of target-based discovery to identify drugs with an ideal activity profile for the prevention, treatment and elimination of malaria.

To eliminate malaria, broadly acting medicines must be developed that cure the symptomatic asexual blood stage, clear the preceding liver-stage infection that can cause relapses and block parasite transmission to mosquitoes¹. Relapse prevention is especially important for *P. vivax*, which forms intra-hepatic hypnozoites that can persist for years before reinitiating development and triggering blood-stage infection. Primaquine is the only licensed antimalarial capable of eliminating the hypnozoite reservoir and delivering a radical cure. However, side effects and weak activity against blood stages preclude widespread use of primaquine². Because the target of primaquine and mechanism of action are not known, the search for radical cure drugs has been limited to related analogues, such as tafenoquine³. There is a clear need for druggable and chemically validated targets that are essential in all life-cycle stages of the malaria parasite.

Here we report that a parasite PI(4)K type III beta (PI(4)KIIIβ), a ubiquitous eukaryotic enzyme that phosphorylates lipids to regulate intracellular signalling and trafficking, is inhibited by imidazopyrazines. In blood stages, imidazopyrazines block a late step in parasite development by disrupting plasma membrane ingression around developing daughter merozoites. This probably stems from altered phosphatidylinositol-4-phosphate (PI(4)P) pools and disrupted Rab11A-mediated membrane trafficking. To our knowledge, these findings are the first validation of a drug target that is required across all *Plasmodium* life-cycle stages.

Imidazopyrazines target several life stages

A cell-based screen against *P. falciparum* asexual blood-stage parasites⁴ revealed a new class of antimalarials with an imidazopyrazine core

(Fig. 1a). This compound was distinct from known antimalarials and was active against several drug-resistant strains (half-maximum inhibitory concentration (IC₅₀) 27–70 nM; Extended Data Table 1 and Extended Data Fig. 1a), suggesting a new mechanism of action. Synthetic derivatives with potency and pharmacokinetic properties suitable for testing in animal models were designed and showed activity *in vivo* (data shown for KDU691; Extended Data Table 2 and Extended Data Fig. 1b).

We evaluated imidazopyrazines (Fig. 1a) against all parasite life-cycle stages in the vertebrate host. First, a cell-based assay showed potent inhibition of liver-stage development of the rodent parasite, *P. yoelii*, with IC₅₀ values less than 160 nM and as low as 9 nM (Fig. 1b) for KDU691, KAI407 and KAI715. *In vivo*, KDU691 administered orally at the time of infection prophylactically protected mice from colonization with transgenic *P. berghei* sporozoites expressing luciferase. Just a single dose (7.5 mg kg⁻¹; Fig. 1c and Extended Data Fig. 1c) prevented what would normally be a fatal outcome. To determine whether KDU691 could eliminate an existing infection, animals were dosed 24, 36 or 48 h after sporozoite inoculation, well after the onset of liver-stage development (which lasts 48 h in *P. berghei*). This delayed treatment rapidly eliminated the luciferase-expressing parasites with a single dose (Fig. 1c). KDU691 was also active at low concentrations (IC₅₀ ~ 196 nM; Fig. 1d) against *in vitro* cultured liver-resident hypnozoites of the simian parasite *P. cynomolgi*⁵. Collectively, these data demonstrate that KDU691 is active against liver stages in several *Plasmodium* species, including hypnozoites associated with malaria relapse. To assess imidazopyrazine activity against a human parasite that forms hypnozoites, we tested

¹Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA. ²Department of Microbiology & Immunology, Columbia University Medical Center, New York, New York 10032, USA. ³Novartis Institutes for Tropical Disease, 138670 Singapore. ⁴Department of Pediatrics, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA. ⁵Department of Parasitology, Biomedical Primate Research Centre, PO Box 3306, 2280 GH Rijswijk, The Netherlands. ⁶TropiQ Health Sciences, 6525 GA Nijmegen, The Netherlands. ⁷Swiss Tropical and Public Health Institute, CH-4002 Basel, Switzerland. ⁸University of Basel, CH-4003 Basel, Switzerland. ⁹Department of Medicine, School of Medicine, Boston University, Boston, Massachusetts 02118, USA. ¹⁰Novartis Institutes for BioMedical Research, CH-4002 Basel, Switzerland. ¹¹Department of Medical Microbiology, Radboud University, Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands. ¹²Laboratory of Malaria Immunobiology, Singapore Immunology Network, Agency for Science Technology and Research (A*STAR), Biopolis, 138648 Singapore. ¹³Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, 117545 Singapore. ¹⁴Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK. ¹⁵Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot 63110, Thailand. ¹⁶Division of Infectious Diseases, Department of Medicine, Columbia University Medical Center, New York, New York 10032, USA.

*These authors contributed equally to this work.

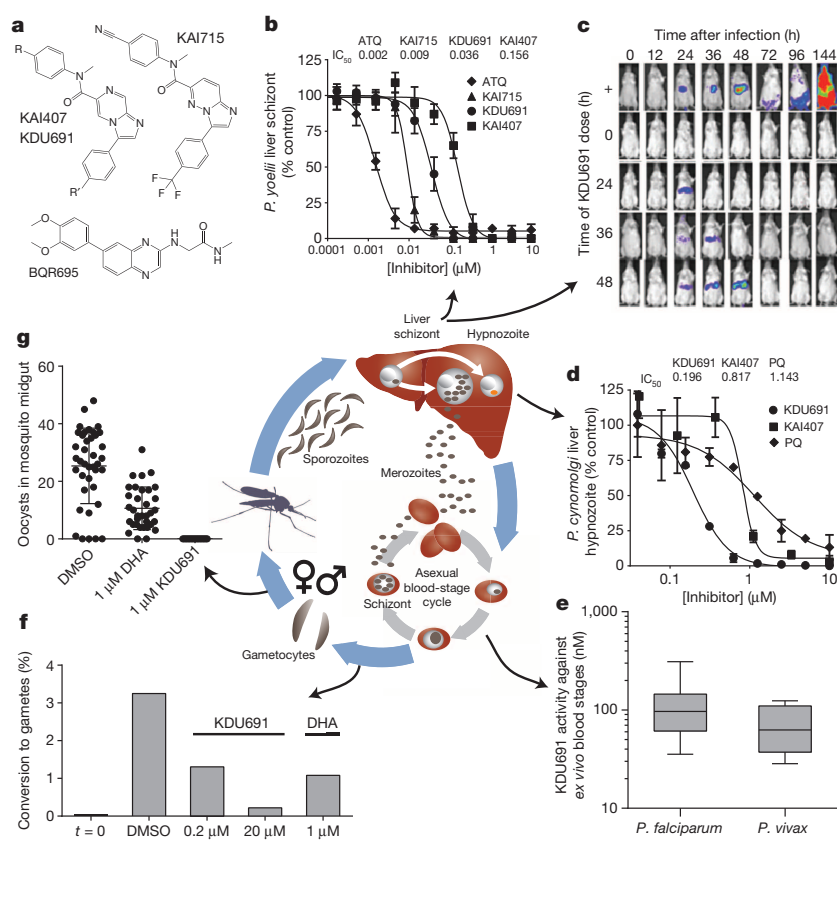


Figure 1 | Imidazopyrazines demonstrate potent activity across *Plasmodium* species and parasite stages. The central schematic illustrates the *Plasmodium* life cycle in mosquitoes and the vertebrate host. Injected sporozoites infect the liver, proliferate and emerge into the bloodstream as merozoites; in some species hypnozoites may remain dormant in the liver. Merozoites undergo several cycles of asexual proliferation in RBCs, or at low frequency, differentiate into sexual-stage gametocytes. Mature gametocytes ingested by an *Anopheles* mosquito become gametes that mate and form oocysts, within which sporozoites develop that migrate to the salivary glands. **a**, Chemical structures of the imidazopyrazine analogues KAI407 (R, trifluoromethyl; R', cyano), KDU691 (R, methylamide; R', chloro) and KAI715, and the quinoxaline BQR695. **b**, *In vitro* activity of imidazopyrazines against liver-stage schizonts of *P. yoelii*. **c**, *In vivo* efficacy of KDU691 against luciferase-expressing *P. berghei*. Mice were either untreated (+) or given a single oral dose of 7.5 mg kg⁻¹ KDU691 before sporozoite inoculation ($t = 0$ h) or after a liver-stage infection was established ($t = 24, 36$ or 48 h after infection), then monitored by bioluminescence (one representative mouse per group shown; $n = 8$). **d**, *In vitro* inhibition of *P. cynomolgi* hypnozoites by imidazopyrazines. **e**, *Ex vivo* analysis of imidazopyrazine activity against asexual blood stages of *P. falciparum* and *P. vivax* field isolates, shown as a box plot (mean, interquartile range 25–75%, and minimum/maximum values). **f**, KDU691 inhibition of *P. falciparum* gametocyte conversion to female gametes after 24 h incubation with compound, expressed as the percentage of Pf525-positive female gametes. **g**, Transmission-blocking effect of KDU691, measured by the number of *P. falciparum* oocysts in *A. stephensi* midguts infected with parasites exposed to 0.1% dimethylsulphoxide (DMSO), 1 μ M DHA or 1 μ M KDU691. ATQ, atovaquone; DHA, dihydroartemisinin; PQ, primaquine. Data are mean \pm s.d. (**b**, **d**), $n = 4$ (**b**, **d**), 8 (*P. falciparum*, **e**), 6 (*P. vivax*, **e**), 2 (**f**) and 20 (**g**).

blood-stage *P. vivax* field isolates for sensitivity. KDU691 potency against *P. vivax* (mean $IC_{50} \sim 69$ nM) was comparable to that observed in parallel studies conducted with field isolates of the more prevalent human pathogen *P. falciparum* (mean $IC_{50} \sim 118$ nM; Fig. 1e).

A small proportion of asexual blood-stage parasites sexually differentiate into gametocytes that transmit to the mosquito vector. Gametocytes are often not killed by drugs that provide symptomatic relief⁶. We sought to determine the transmission-blocking potential by assessing compound activity on gametocyte viability and the mosquito-specific stages of gamete and oocyst formation. An enriched gametocyte population was treated with either KDU691 or dihydroartemisinin (DHA), a licensed antimalarial with gametocytocidal activity⁷. Both compounds reduced gametocyte viability in a dose-dependent manner, with KDU691 displaying an IC_{50} value of 220 nM (Extended Data Fig. 1d). Furthermore, 200 nM KDU691 reduced gamete formation by 60%, comparable to that achieved with 1 μ M DHA (Fig. 1f). Lastly, we tested compounds in the 'standard membrane feeding assay' that directly monitors transmission to the mosquito⁸. Transmission was completely inhibited at 1 μ M KDU691 (Fig. 1g), with dose-dependent suppression of oocyst densities ($IC_{50} \sim 316$ nM; Extended Data Fig. 1e) and oocyst prevalence (that is, the number of infected mosquitoes; $IC_{50} \sim 370$ nM; Extended Data Fig. 1f). Altogether, these data indicate that KDU691 is effective at blocking transmission, and suggest it targets a conserved function throughout the *Plasmodium* life cycle.

Inhibition of blood-stage cytokinesis

Imidazopyrazine-treated parasites accumulated multinucleated schizonts (Fig. 2a), leading us to test compound effects when administered at distinct stages during development in red blood cells (RBCs). Parasite reinvasion was markedly inhibited even when KAI407 was added late in the intraerythrocytic cycle, up to 44 h after invasion (Fig. 2b). Conversely, early ring-stage parasites exposed continuously to KAI407

for up to 40 h followed by drug washout demonstrated no appreciable loss of parasite viability (Extended Data Fig. 2a). These data indicate that the imidazopyrazine target is most critical at a very late stage in the 48 h intraerythrocytic cycle.

We next tested whether drug-treated parasites arrest during maturation of daughter merozoites or form daughter cells that fail to egress from the host cell^{9,10}. Synchronized imidazopyrazine-treated parasites were mechanically ruptured, yielding a reinvasion parasitaemia near 0.01% (Fig. 2c), suggesting a defect in merozoite formation. As a control, merozoites that were artificially retained within a mature parasite by treatment with E-64, a protease inhibitor that blocks egress¹¹, and then mechanically released remained viable and capable of reinfecting RBCs, yielding 2–3% parasitaemia^{9,10} (Fig. 2c). Mock-treated parasites that had been allowed to invade naturally yielded a parasitaemia of 8–10%.

To visualize drug-induced cellular defects, we examined a transgenic line expressing green fluorescent protein (GFP)-tagged PfATP4, a plasma-membrane protein¹² that reveals merozoite envelopment during biogenesis. Control E-64-treated parasites showed complete plasma membrane ingression around each daughter cell (Fig. 2d and Extended Data Fig. 2b). By contrast, KAI407-treated parasites showed incomplete and disorganized segmentation, lacking merozoites that were fully enclosed by membranes with PfATP4-GFP, suggesting perturbation of plasma membrane ingression (Fig. 2d and Extended Data Fig. 2c). On the basis of the timing of growth inhibition and the observed cellular phenotype, we conclude that imidazopyrazines impair cytokinesis by interfering with membrane biogenesis around the developing merozoite.

Plasmodium PI(4)K target identification

To determine the mechanism of broad antimalarial activity by imidazopyrazines, we next sought to elucidate their target using forward genetics. We selected imidazopyrazine-resistant blood-stage *P. falciparum*

parasites *in vitro* using either KAI407 or KAI715. Resistant lines were evolved in three independent flasks for each compound, with several clones isolated from each flask for IC₅₀ characterization (Fig. 3a). For each clone, we identified genomic differences relative to the parent using a genome-tiling microarray¹³. All five KAI407-resistant clones shared alterations to a single gene, *pfpi4k* (PlasmoDB accession PF3D7_0509800; PI(4)KIIIβ), one of two annotated PI(4)Ks. Type IIIβ phosphatidylinositide kinases (Extended Data Fig. 3a, b) are conserved among eukaryotes and function to convert phosphatidylinositol to PI(4)P. Within the cell, PI(4)P regulates effector protein recruitment and lipid-sorting events at the Golgi, and is a precursor for the generation of phosphatidylinositol-4,5-bisphosphate at the plasma membrane^{14,15}.

Three KAI407-resistant clones possessed copy number variations (CNVs) that included *pfpi4k* (Extended Data Fig. 4a). In the remaining two clones, a single nucleotide variant (SNV) was detected in the second exon of *pfpi4k* (Extended Data Fig. 4b). Sequencing of *pfpi4k* in these two independently evolved clones revealed an identical non-synonymous SNV that led to a His1484Tyr coding change. Removal of drug pressure for 3 months resulted in resensitization of one clone (KAI407-R3-REV) to KAI407, with a corresponding reversion in *pfpi4k* to the wild-type sequence, supporting a role for this variant in KAI407 resistance (Fig. 3a). Selections with KAI715 yielded two resistant clones, one of which (KAI715-R4) also implicated *pfpi4k* amplification in resistance (Extended Data Fig. 4c). The second clone (KAI715-R5) possessed a solitary Asp139Tyr mutation in a second gene product, the small GTPase Pfrab11A (PF3D7_1320600) (Extended Data Fig. 4d). Notably, PI(4)K binds and regulates Rab11A in *Drosophila* spermatocytes¹⁶, consistent with a related function in membrane trafficking.

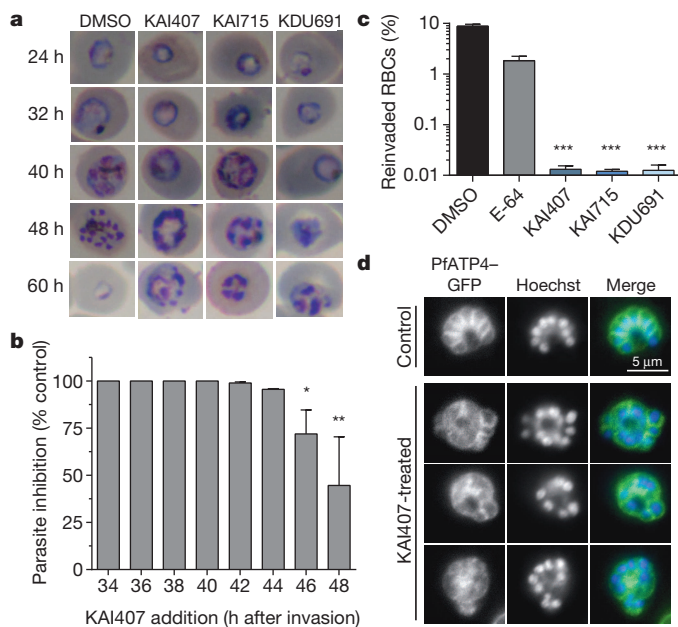


Figure 2 | PfPI(4)K is required for the completion of cytokinesis in asexual blood-stage schizonts. **a**, Microscopy of highly synchronized parasites continuously treated with $\sim 5 \times \text{IC}_{50}$ drug (125 nM KAI407, 15 nM KAI715 or 150 nM KDU691) or DMSO vehicle. Representative images from a single experimental replicate are shown. Original magnification, $\times 100$. **b**, Drug-free media was supplemented with 125 nM KAI407 in 2-h intervals starting at the mid/late trophozoite stage ($t = 34$ h after invasion). Parasitaemia in the next life cycle ($t = 72$ h) was normalized to untreated parasites. **c**, Percentage reinvasion of RBCs after mechanical rupture of schizonts treated with 125 nM KAI407, 15 nM KAI715, 150 nM KDU691 or 1 μM E-64. **d**, PfATP4-GFP was used to visualize plasma membrane ingression around developing daughter merozoites, with nuclei stained by Hoechst 33342. Representative images from a single experimental replicate are shown. Scale bar, 5 μm. Parasites treated with 500 nM KAI407 for 4 h had a disorganized membrane structure, without clearly defined daughter cells. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ (Mann-Whitney U test). Data are mean \pm s.d. (**b**, **c**); $n = 3$ (**a**), 4 (**b**, **c**) and 2 (**d**).

Our imidazopyrazine-resistant lines provide compelling evidence for PI(4)KIIIβ as the plasmodial drug target. Orthologues of PI(4)KIIIβ are found in all *Plasmodium* species and are extensively conserved at the amino acid level, including 97% identity in the catalytic domain between the *P. falciparum* and *P. vivax* orthologues (Extended Data Fig. 3b). Conservation with the catalytic region of human PI(4)KIIIβ is less significant, with only 43% identity. The partial sequence conservation between the human and *Plasmodium* orthologues of PI(4)KIIIβ (Extended Data Fig. 3b) led us to test whether some inhibitors of the human kinase might also exert antimalarial activity.

Subsequent screening of a Novartis compound library led us to identify BQR695 (Fig. 1a), a quinoxaline that displayed sub-micromolar potency against both human PI(4)KIIIβ (IC₅₀ ~ 90 nM; Extended Data Table 3) and *P. falciparum* asexual blood stages (IC₅₀ ~ 71 nM). BQR695, which showed no evidence of toxicity against mature RBCs, induced a schizont-stage arrest indistinguishable from that observed in imidazopyrazine-treated parasites (Extended Data Fig. 2d, e), and exhibited cross-resistance with the imidazopyrazine-resistant lines (Fig. 3a). These data confirm a shared mechanism of action and of resistance with the imidazopyrazines. To test this association further, we evolved BQR695-resistant parasites. Whole-genome scanning of three independently selected lines (BQR695-R) showed two instances of CNVs involving *pfpi4k* (Extended Data Fig. 4e), and a new mutation (Ser1320Leu) in *pfpi4k* (Extended Data Fig. 4f). Sequencing of *pfpi4k* in the CNV-containing lines revealed that BQR695-R7 had at least one *pfpi4k* copy with a Tyr1356Phe coding change, which was associated with a higher degree of resistance to BQR695 (Fig. 3a). All three resistance-associated SNVs, Ser1320Leu, Tyr1356Phe and His1484Tyr, mapped to the catalytic domain of PfPI(4)K (Extended Data Fig. 3a, b). Notably, the three imidazopyrazines (KAI407, KAI715 and KDU691) and the quinoxaline BQR695 demonstrated comparable cross-resistance against all drug-resistant lines (Fig. 3a). Collectively, these data provide compelling chemical and genetic evidence that PI(4)K is the direct drug target of imidazopyrazine and quinoxaline compounds.

PI(4)K variation confers drug resistance

To confirm the role of *pfpi4k* and *pfrab11a* mutations in resistance, we engineered these variants into drug-sensitive parasites. We generated recombinant parasites bearing individual *pfpi4k* SNVs using zinc-finger nuclease (ZFN) technology¹⁷. Genome editing was confirmed by direct sequencing (Fig. 3b) and whole-genome analysis ($>20\times$ coverage across 96.4% of the genome; Extended Data Table 4). Genome-edited Dd2 parasites containing either the Ser1320Leu or His1484Tyr mutations were resistant to both imidazopyrazine and quinoxaline compounds (Fig. 3c and Extended Data Fig. 5a) and yielded resistance comparable to the original drug-evolved lines. By contrast, the Tyr1356Phe mutation conferred resistance specifically to the quinoxaline BQR695 but not to the imidazopyrazines (Fig. 3c). In comparison, the evolved line BQR695-R7, which possessed both the Tyr1356Phe mutation as well as a CNV, showed resistance to both imidazopyrazine and quinoxaline compounds (Fig. 3a).

Because several resistant lines contained CNV events, we sought to determine whether increased copy numbers of wild-type PI(4)K could alone mediate protection. We generated a stable parasite line over-expressing GFP-tagged PI(4)K (ref. 18). GFP-PI(4)K was diffusely localized in trophozoite-stage parasites and enriched at the apical ends of developing daughter merozoites (Extended Data Fig. 3c). Expression of this additional wild-type copy of PfPI(4)K resulted in decreased susceptibility to both imidazopyrazine and quinoxaline compounds, confirming that CNVs can contribute to resistance and providing evidence that the GFP-tagged protein is functional (Fig. 3d). Collectively, these data demonstrate that PfPI(4)K can mediate resistance to the imidazopyrazines in asexual blood stages through either increased copy number or point mutations.

We next investigated whether PI(4)K variants could confer resistance in liver-stage parasites. Introduction of a mutation analogous to

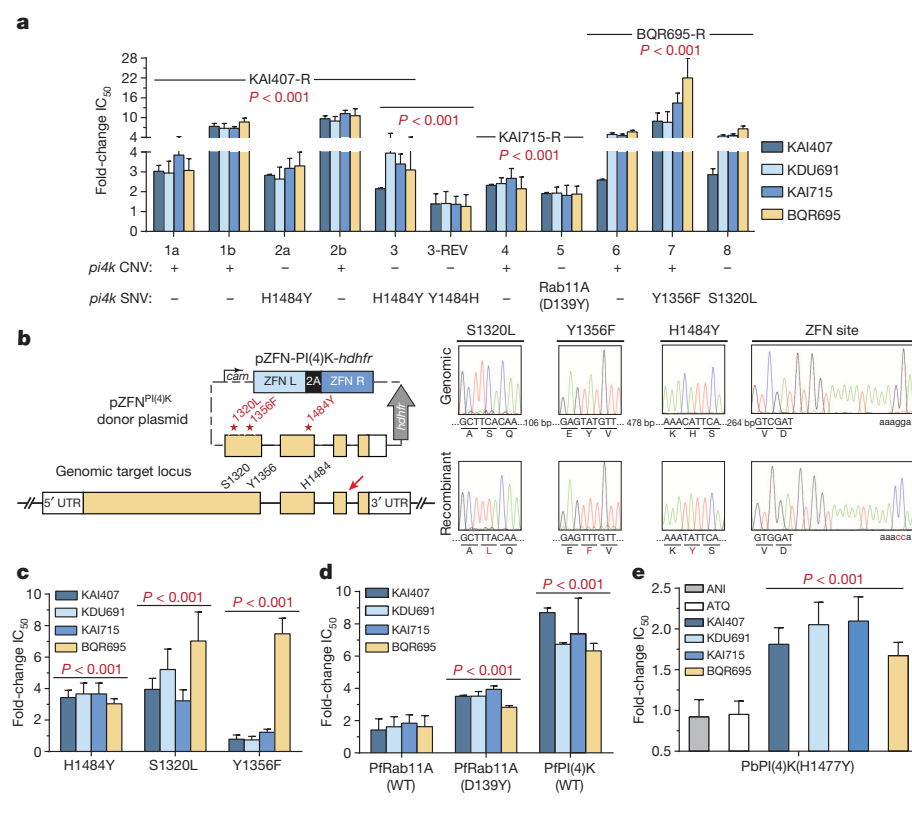


Figure 3 | Resistance to imidazopyrazine and quinoxaline compounds is mediated by gene variations in PfPI(4)K and PfRab11A. **a**, Fold change in IC_{50} values between 11 drug-evolved clones and the drug-sensitive parent, for the imidazopyrazines KAI407, KAI715 and KDU691, and the quinoxaline BQR695. Statistically significant mean IC_{50} values for each drug-resistant line were identified using the Mann–Whitney U test (also for **c–e**). CNVs and SNVs are noted. **b**, Schematic overview of the *pfp4k* gene editing strategy to introduce putative resistance mutations into a wild-type parasite. ZFNs target a 34-base-pair (bp) site on *pfp4k* (red arrow); after cleavage, homology-dependent repair from a 1.7-kilobase (kb) donor sequence resulted in incorporation of the specific SNV (shown in red), as well as additional silent mutations at the ZFN cut site. **c**, Fold change in IC_{50} values of ZFN-edited lines of the PfPI(4)K mutants His1484Tyr, Ser1320Leu and Tyr1356Phe against KAI407, KDU691, KAI715 and BQR695. IC_{50} values for other known antimalarials are provided in Extended Data Fig. 5a. **d**, Chromosomal integration of GFP fusions of wild-type Rab11A, mutant Rab11A(Asp139Tyr) or wild-type PI(4)K conferred resistance to imidazopyrazine and quinoxaline compounds. A control line expressing GFP alone was used to normalize the IC_{50} shift. **e**, Fold change in IC_{50} values of *P. berghei* *in vitro* liver-stage schizonts expressing PbPI(4)K(His1477Yr) (equivalent to PfPI(4)K(His1484Tyr)). Data are mean \pm s.d. (**a**, **c–e**; $n = 8$ (**a**), 12 (**c**), 6 (**d**) and 12 (**e**)).

His1484Tyr into *P. berghei* (His1477Tyr) conferred a significant IC_{50} shift for the imidazopyrazine and quinoxaline compounds against *in vitro* liver-stage parasites (Fig. 3e). Importantly, these data show that the mechanism of action for these inhibitors is conserved between life-cycle stages.

We next tested the validity of the Rab11A(Asp139Tyr) variation, which conferred a modest, but significant, resistance phenotype (KAI715-R5; Fig. 3a). The Asp139Tyr SNV lies on the α 4 helix of Rab11A, distant from the GTP-binding site (Extended Data Fig. 6a, b). GFP-tagged versions of both wild-type and Rab11A(Asp139Tyr) showed a predominantly cytosolic localization that became apically punctate in late schizonts (Extended Data Fig. 6c, d), consistent with a previous report¹⁹. As expected, the Asp139Tyr mutant conferred modest resistance against KAI407 (Fig. 3d).

If PI(4)K is the cellular target of the imidazopyrazines, then its function should be essential in most stages of the *Plasmodium* life cycle. We therefore tested the essentiality of *pfp4k* in the asexual blood stages by using ZFNs to introduce a stop codon at position Tyr 1356, which lies within the catalytic site of the kinase domain (see Fig. 4a, b). This Tyr 1356 stop-editing experiment was attempted in three independent transfections. Unlike the Tyr1356Phe mutation, which was readily isolated in $\sim 18\%$ of clones, two transfection reactions produced 0 out of 35 clones with a stop codon. From a third transfection, one clone yielded a duplication on chromosome 5 that partially encompassed *pfp4k* and placed the edited stop codon within a truncated copy of the gene (Extended Data Fig. 5b, c). These data strongly indicate that gene editing could not be achieved without a duplication event that preserved one copy of the gene, confirming *pfp4k* essentiality to blood-stage growth.

Biochemical validation of PfPI(4)K

We performed *in silico* modelling on a PfPI(4)K homology model based on related phosphatidylinositol-3-OH kinases (Fig. 4a), finding that both KAI407 and BQR695 were accommodated in the ATP-binding pocket. The nitrogen at position 1 of the imidazole ring of KAI407 is predicted to make a crucial hydrogen bond contact with the main-chain amide from Val 1357 that lies adjacent to the Tyr1356Phe mutation

(Fig. 4b). This ‘hinge-binding’ contact is common among protein kinase inhibitors, and mimics the hydrogen bonds made by the adenine of ATP²⁰.

To confirm that imidazopyrazines directly inhibit PI(4)K function, we expressed full-length recombinant *P. vivax* PI(4)K (PvPI(4)K). KDU691 and BQR695 showed dose-dependent inhibition of PvPI(4)K activity (Fig. 4c, d), with IC_{50} values of 1.5 and 3.5 nM, respectively, in the presence of 10 μ M ATP. Furthermore, compound potency decreased with increasing concentrations of ATP, indicative of ATP-competitive inhibition, providing further evidence that imidazopyrazines and quinoxalines bind to the ATP-binding pocket of parasite PI(4)K. Mutations homologous to the PfPI(4)K resistance-conferring variants (Extended Data Fig. 3b) generated resistance profiles (Fig. 4e) similar to those observed with the respective drug-resistant parasite lines (Fig. 3a, c). In contrast to the low nanomolar activity observed against parasite PI(4)K, KDU691 and KAI407 did not significantly inhibit any of the >40 enzymes in a panel of human kinases, and both were at least 1,000 times more potent against the parasite enzyme than against any human lipid kinase (Extended Data Table 3), demonstrating remarkable selectivity. This specificity precludes the possibility that imidazopyrazine antimalarial activity derives from the inhibition of host kinases.

Perturbed cellular distribution of PI(4)P

Finally, we examined the consequences of PfPI(4)K inhibition on intracellular PI(4)P. Unexpectedly, high-performance liquid chromatography (HPLC) analysis of radiolabelled phospholipids did not reveal a decrease in total PI(4)P after imidazopyrazine treatment (Extended Data Fig. 3d). To determine whether imidazopyrazines might instead change the intracellular distribution of PI(4)P, we generated a transgenic parasite expressing a validated PI(4)P-specific probe, GFP–PH^{Osh2}, which consists of GFP fused to two tandem PI(4)P-binding pleckstrin homology (PH) domains from yeast Osh2 (ref. 21) (Fig. 4f). In yeast, this marker defines two populations of PI(4)P—one at the plasma membrane, and one intracellular population, with the latter diminished by inhibition of the Golgi-localized PI(4)K homologue Pik1 (ref. 21). Similarly, in *P. falciparum*, GFP–PH^{Osh2} localized to intracellular foci, and, to a lesser degree, the parasite plasma

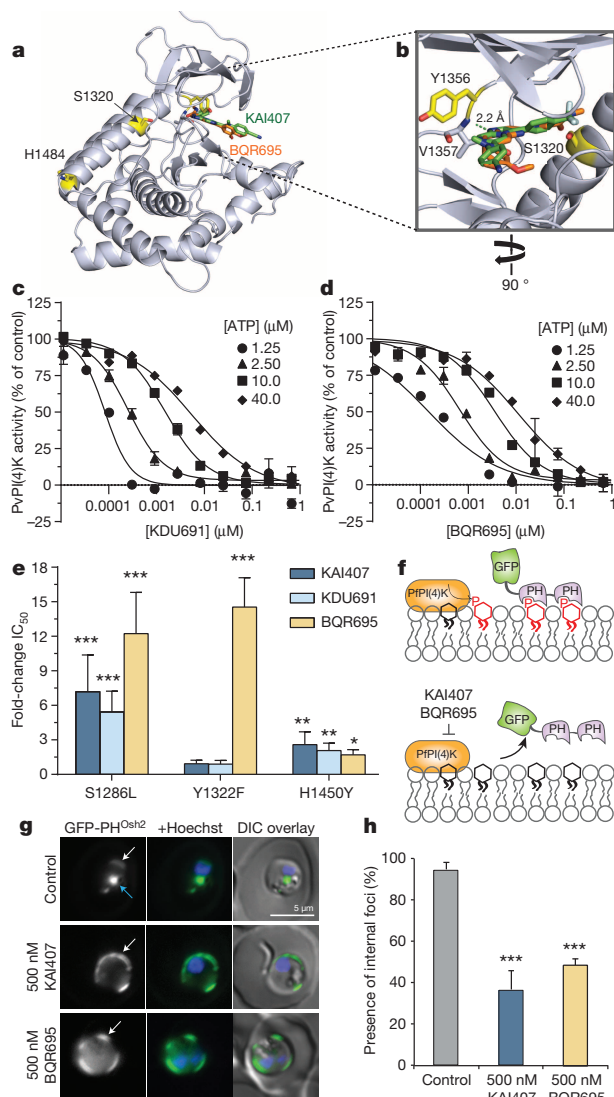


Figure 4 | Imidazopyrazines and quinoxalines bind the ATP-binding site of PI(4)K and alter PI(4)P intracellular distribution. **a**, *In silico* docking of a homology model of the PfPI(4)K catalytic domain showing KAI407 (green) and BQR695 (orange) binding to the ATP-binding pocket of PfPI(4)K (cartoon representation; grey). Resistance mutations (yellow) are labelled. **b**, Zoomed view of the ATP-binding pocket rotated 90° along the y axis, showing the predicted hydrogen bond (green dash) between the imidazole nitrogen in KAI407 and the amide bond from the Val 1357 backbone (grey sticks). **c**, **d**, Recombinant full-length *P. vivax* PI(4)K (PvPI(4)K) activity in the presence of KDU691 (**c**) or BQR695 (**d**), monitored across a range of ATP concentrations (1.25–40 μM). **e**, Recombinant PvPI(4)K mutants Ser1286Leu, Tyr1322Phe or His1450Tyr (equivalent to PfPI(4)K mutants Ser1320Leu, Tyr1356Phe or His1484Tyr, respectively) were assayed against imidazopyrazine and quinoxaline compounds. Fold change in IC₅₀ value is shown relative to wild-type enzyme. **f**, Illustration of the PI(4)P-dependent localization of GFP-PH^{Osh2}. **g**, *In vivo* distribution of PI(4)P was detected in parasites expressing GFP-PH^{Osh2}. In untreated parasites, GFP-PH^{Osh2} localized to intracellular foci (blue arrow) and the plasma membrane (white arrows). Treatment with 500 nM KAI407 or BQR695 for 4 h depleted the intracellular pool and redistributed the probe to the plasma membrane. Representative images from a single experimental replicate are shown. Nuclei were stained with Hoechst 33342 (blue); DIC, differential interference contrast. Scale bar, 5 μm. **h**, Quantification of intracellular GFP-PH^{Osh2}-labelled foci after drug treatment. **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (Mann–Whitney *U* test). Data are mean ± s.d. (**c–e**, **h**); *n* = 9 (**c–e**) and 3 (**g**, **h**).

membrane (Fig. 4g). Treatment with 0.5 μM of either KAI407 or BQR695 caused GFP-PH^{Osh2} to redistribute to the parasite plasma membrane (Fig. 4g, h), consistent with depletion of intracellular PI(4)P after inhibition of PfPI(4)K function.

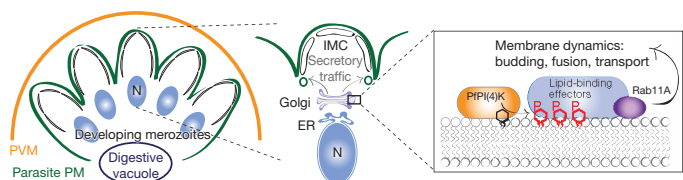


Figure 5 | Proposed role of PfPI(4)K within asexual blood-stage parasites. Model depicting daughter merozoite biogenesis in a late-stage schizont, showing plasma membrane (PM) ingression and development of the underlying inner membrane complex (IMC). Delivery of post-Golgi secretory vesicles to the invaginating plasma membrane requires PfPI(4)K-driven production of PI(4)P, which probably recruits lipid-binding effector proteins in a Rab11A-dependent process. N, nucleus; PVM, parasitophorous vacuolar membrane.

Discussion

Here we report the identification of PI(4)K as an essential and drugable target of imidazopyrazine and quinoxaline compounds that block development at multiple stages of the *Plasmodium* life cycle, implicating this phosphatidylinositol kinase as a crucial cellular regulatory molecule. PfPI(4)K functionally complements yeast Pik1, an essential Golgi-localized PI(4)KIIIβ (ref. 22). Our data indicate that, like Pik1 (refs 23–25), PfPI(4)K regulates essential membrane trafficking events, with the plasmodial protein required for the late-stage membrane ingression that drives completion of the asexual erythrocytic stages. We propose that *Plasmodium* PI(4)K acts at the Golgi to locally generate PI(4)P, which in turn recruits effectors that generate transport vesicles destined for the ingressing plasma membrane (Fig. 5). One obvious candidate effector is Rab11A, which has also been linked to PI(4)K in *Drosophila* and mammalian cells^{16,26}. *Drosophila* PI(4)K and Rab11 act during cytokinesis, regulating membrane delivery to complete cell division¹⁶, which mirrors our findings of failed membrane ingression after PfPI(4)K inhibition. Rab11A may bind directly to PI(4)K and/or PI(4)P, or may instead be recruited indirectly by an intermediate protein. In either case, Rab11A probably marks these Golgi-derived vesicles to modulate delivery to the target compartment, a role consistent with previous findings implicating *Plasmodium* Rab11A in cytokinesis¹⁹.

Our results provide genetic, chemical and biochemical validation of *Plasmodium* PI(4)K as a target suitable for antimalarial drug discovery. Although the imidazopyrazine compounds described here may serve primarily as tool compounds, further optimization could lead to clinical candidates with desirable drug-like characteristics. Maintaining their remarkable selectivity for *Plasmodium* PI(4)K will be an important feature for an ultimate antimalarial drug because chronic PI(4)KIIIβ inhibition in humans causes immunosuppressive effects²⁷ that could preclude long-term malaria prophylaxis. Because the emergence of resistant parasites has diminished the effectiveness of all licensed antimalarial therapies, most of which were introduced as monotherapies, clinical candidates will necessarily require combination with mechanistically distinct drug partner(s) to mitigate the potential for drug resistance and to complement the PI(4)K inhibitors' narrow and late stage of action in the asexual blood stage.

Collectively, our data identify *Plasmodium* PI(4)K as a new drug target that operates in all life-cycle stages of the parasite, and is therefore suitable for the development of next-generation antimalarial drugs capable of curing, preventing and blocking the transmission of the disease. We anticipate that our findings will rapidly yield clinical candidates compatible with single exposure, radical cure and prophylaxis, a profile widely heralded as crucial for the success of worldwide malaria elimination efforts²⁸.

METHODS SUMMARY

Imidazopyrazine and quinoxaline activity were demonstrated throughout the vertebrate life cycle, with eight specific assays detailed in the Methods: *in vitro* high-content imaging assay against *P. yoelii*²⁹ and *P. cynomolgi* exo-erythrocytic forms; *in vivo* causal prophylaxis mouse model, in which parasitaemia was monitored by *in vivo* imaging of the luciferase-expressing *P. berghei* line²⁹; *in vitro* asexual-stage

activity against *P. falciparum*, measured using a modified SYBR Green cell proliferation assay⁴; *ex vivo* asexual-stage assay against *P. vivax* and *P. falciparum* isolates collected from Thailand^{30,31}; gametocyte viability assays against *P. falciparum* NF54 stage III–V gametocytes measured by a modified *Plasmodium* lactate dehydrogenase assay³²; *in vitro* gametogenesis assay against *P. falciparum* quantified by flow cytometry³³; and *P. falciparum* transmission to *Anopheles stephensi* mosquitoes in a standard membrane feeding assay³⁴. Stage of action studies¹² and a merozoite release assay^{9,10} were used to identify specific arrest of drug-treated asexual parasites in mid-schizogony. An established protocol¹³ was used for *in vitro* evolution and comparative genomic analysis of KAI407-, KAI715- or BQR695-resistant lines. Resistance-conferring mutations were validated with ZFN approaches¹⁷ and standard transfection-based methods¹². Genomic DNA libraries generated on an Illumina HiSeq 2000 were analysed with PlatypUS, an internal WGS pipeline. A *P. berghei* line expressing PI(4)K(His1477Tyr) was generated by standard methods³⁵, and resistance was quantified against liver-stage parasites²⁹. A homology model of PfPI(4)K was built from atomic coordinates for human phosphatidylinositol-3-OH kinase gamma (PI(3)K γ ; Protein Data Bank accession 3ENE) and used for flexible ligand docking studies within Glide (Schrödinger). Full-length PvPI(4)K was expressed in a SF9/baculovirus system and assayed for drug inhibition with the Transcreeper ADP₂ FP detection kit (BellBrook). PI(4)P pools were detected in asexual stage parasites by expression of tandem Osh2 pleckstrin homology domains fused to GFP²¹.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July; accepted 15 October 2013.

Published online 27 November 2013.

- Greenwood, B. M. *et al.* Malaria: progress, perils, and prospects for eradication. *J. Clin. Invest.* **118**, 1266–1276 (2008).
- Vale, N., Moreira, R. & Gomes, P. Primaquine revisited six decades after its discovery. *Eur. J. Med. Chem.* **44**, 937–953 (2009).
- Wells, T. N., Burrows, J. N. & Baird, J. K. Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination. *Trends Parasitol.* **26**, 145–151 (2010).
- Plouffe, D. *et al.* *In silico* activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proc. Natl Acad. Sci. USA* **105**, 9059–9064 (2008).
- Dembelle, L. *et al.* Towards an *in vitro* model of *Plasmodium* hypnozoites suitable for drug discovery. *PLoS ONE* **6**, e18162 (2011).
- Bousema, T. & Drakeley, C. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. *Clin. Microbiol. Rev.* **24**, 377–410 (2011).
- Adjalley, S. H. *et al.* Quantitative assessment of *Plasmodium falciparum* sexual development reveals potent transmission-blocking activity by methylene blue. *Proc. Natl Acad. Sci. USA* **108**, E1214–E1223 (2011).
- van Pelt-Koops, J. C. *et al.* The spiroindolone drug candidate NITD609 potentially inhibits gametocytogenesis and blocks *Plasmodium falciparum* transmission to *Anopheles* mosquito vector. *Antimicrob. Agents Chemother.* **56**, 3544–3548 (2012).
- Boyle, M. J. *et al.* Isolation of viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and advance vaccine and drug development. *Proc. Natl Acad. Sci. USA* **107**, 14378–14383 (2010).
- Dvorin, J. D. *et al.* A plant-like kinase in *Plasmodium falciparum* regulates parasite egress from erythrocytes. *Science* **328**, 910–912 (2010).
- Drew, M. E. *et al.* *Plasmodium* food vacuole plasmepsins are activated by falcipains. *J. Biol. Chem.* **283**, 12870–12876 (2008).
- Rottmann, M. *et al.* Spiroindolones, a potent compound class for the treatment of malaria. *Science* **329**, 1175–1180 (2010).
- Dharia, N. V. *et al.* Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol.* **10**, R21 (2009).
- Balla, A. & Balla, T. Phosphatidylinositol 4-kinases: old enzymes with emerging functions. *Trends Cell Biol.* **16**, 351–361 (2006).
- Mayinger, P. Phosphoinositides and vesicular membrane traffic. *Biochim. Biophys. Acta* **1821**, 1104–1113 (2012).
- Polevoy, G. *et al.* Dual roles for the *Drosophila* PI 4-kinase four wheel drive in localizing Rab11 during cytokinesis. *J. Cell Biol.* **187**, 847–858 (2009).
- Straimer, J. *et al.* Site-specific genome editing in *Plasmodium falciparum* using engineered zinc-finger nucleases. *Nature Methods* **9**, 993–998 (2012).
- Nkrumah, L. J. *et al.* Efficient site-specific integration in *Plasmodium falciparum* chromosomes mediated by mycobacteriophage Bxb1 integrase. *Nature Methods* **3**, 615–621 (2006).
- Agop-Nersesyan, C. *et al.* Rab11A-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis. *PLoS Pathog.* **5**, e1000270 (2009).
- Noble, M. E. M., Endicott, J. A. & Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **303**, 1800–1805 (2004).
- Roy, A. & Levine, T. P. Multiple pools of phosphatidylinositol 4-phosphate detected using the pleckstrin homology domain of Osh2p. *J. Biol. Chem.* **279**, 44683–44689 (2004).
- Krüger, T., Sanchez, C. P. & Lanzer, M. Complementation of *Saccharomyces cerevisiae* *pik1^{ts}* by a phosphatidylinositol 4-kinase from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **172**, 149–151 (2010).
- Strahl, T., Hama, H., DeWald, D. B. & Thorner, J. Yeast phosphatidylinositol 4-kinase, *Pik1*, has essential roles at the Golgi and in the nucleus. *J. Cell Biol.* **171**, 967–979 (2005).
- Walch-Solimena, C. & Novick, P. The yeast phosphatidylinositol-4-OH kinase *Pik1* regulates secretion at the Golgi. *Nature Cell Biol.* **1**, 523–525 (1999).
- Hama, H., Schnieders, E. A., Thorner, J., Takemoto, J. Y. & DeWald, D. B. Direct involvement of phosphatidylinositol 4-phosphate in secretion in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 34294–34300 (1999).
- de Graaf, P. *et al.* Phosphatidylinositol 4-kinase β is critical for functional association of Rab11 with the Golgi complex. *Mol. Biol. Cell* **15**, 2038–2047 (2004).
- LaMarche, M. J. *et al.* Anti-hepatitis C virus activity and toxicity of type III phosphatidylinositol-4-kinase β inhibitors. *Antimicrob. Agents Chemother.* **56**, 5149–5156 (2012).
- The malERA Consultative Group on Drugs. A research agenda for malaria eradication: drugs. *PLoS Med.* **8**, e1000402 (2011).
- Meister, S. *et al.* Imaging of *Plasmodium* liver stages to drive next-generation antimalarial drug discovery. *Science* **334**, 1372–1377 (2011).
- Russell, B. *et al.* Determinants of *in vitro* drug susceptibility testing of *Plasmodium vivax*. *Antimicrob. Agents Chemother.* **52**, 1040–1045 (2008).
- Russell, B. M. *et al.* Simple *in vitro* assay for determining the sensitivity of *Plasmodium vivax* isolates from fresh human blood to antimalarials in areas where *P. vivax* is endemic. *Antimicrob. Agents Chemother.* **47**, 170–173 (2003).
- D'Alessandro, S. *et al.* A *Plasmodium falciparum* screening assay for anti-gametocyte drugs based on parasite lactate dehydrogenase detection. *J. Antimicrob. Chemother.* **68**, 2048–2058 (2013).
- Decherer, K. J., Thompson, J., Dodemont, H. J., Eling, W. & Konings, R. N. Developmentally regulated expression of pfs16, a marker for sexual differentiation of the human malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **89**, 235–244 (1997).
- van der Kolk, M. *et al.* Evaluation of the standard membrane feeding assay (SMFA) for the determination of malaria transmission-reducing activity using empirical data. *Parasitology* **130**, 13–22 (2005).
- Janse, C. J., Ramesar, J. & Waters, A. P. High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei*. *Nature Protocols* **1**, 346–356 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Miller for critical review of the manuscript and figure design. We also thank A. Rodriguez and the insectary core facility team at New York University for reliable supplies of malaria-infected mosquitoes. We gratefully acknowledge translational research grants (WT078285 and WT096157) from the Wellcome Trust and funding from the Medicines for Malaria Venture (MMV) to the Genomics Institute of the Novartis Research Foundation, the Swiss Tropical and Public Health Institute, Columbia University, the Novartis Institute for Tropical Diseases, the Singapore Immunology Network and Horizontal Programme on Infectious Diseases under the Agency Science Technology and Research (A*STAR, Singapore), and the Wellcome Trust (UK). Shoklo Malaria Research Unit is sponsored by The Wellcome Trust (UK), as part of the Oxford Tropical Medicine Research Programme of Wellcome Trust-Mahidol University. E.A.W. and D.A.F. are supported by grants from the Bill and Melinda Gates Foundation, MMV, and the National Institutes of Health (R01AI090141 to E.A.W. and R01085584 and R01079709 to D.A.F.).

Author Contributions C.W.M., with assistance from S.L.M., M.I. and D.M.P., evolved and characterized drug-resistant parasite lines, analysed microarray data, and performed phenotypic studies. M.C.S.L. performed genome editing and other transgenic parasite studies, as well as the fluorescence microscopy imaging. Additional experimental contributions were as follows: PvPI(4)K assay (C.S.L., S.H.L. and C.B.); imidazopyrazine chemistry (J.R., A.N., A.K.C. and D.C.T.); imidazopyrazine structure–activity studies (K.G. and K.L.K.); BQR695 re-synthesis (O.S.); quinoxaline development and human kinase panels (J.T. and D.H.); mutant *P. berghei* strain generation (T.R.S.K. and P.P.H.); next-generation sequencing data analysis (M.J.M.); *in vitro* assay with *P. cynomolgi* (A.-M.Z. and C.H.M.K.); sexual-stage *P. falciparum* assays (M.T., K.J.D. and R.W.S.); *ex vivo* assays on *P. falciparum* and *P. vivax* clinical isolates (R.S., B.R., L.R. and F.N.); *in vivo* efficacy studies in the mouse model (blood stage, C.F. and M.R.; liver stage, N.K.); *in vitro* assay with *P. yoelii* (D.M.P., S.M., S.L.M. and K.G.); *in silico* docking studies (B.B.). L.R. quantified phosphatidylinositol phosphates. R.J.G. managed Genomics Institute of the Novartis Research Foundation (GNF) activities. B.K.S.Y. coordinated collaborative efforts. C.W.M., M.C.S.L., C.B., D.A.F., T.T.D. and E.A.W. designed experiments and co-wrote the manuscript. C.W.M. and M.C.S.L. contributed equally to the study. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A.W. (ewinzeler@ucsd.edu) or T.T.D. (thierry.diagana@novartis.com).

Structural mechanism of ligand activation in human GABA_B receptor

Yong Geng¹, Martin Bush¹, Lidia Mosyak¹, Feng Wang¹ & Qing R. Fan^{1,2}

Human GABA_B (γ -aminobutyric acid class B) receptor is a G-protein-coupled receptor central to inhibitory neurotransmission in the brain. It functions as an obligatory heterodimer of the subunits GBR1 and GBR2. Here we present the crystal structures of a heterodimeric complex between the extracellular domains of GBR1 and GBR2 in the apo, agonist-bound and antagonist-bound forms. The apo and antagonist-bound structures represent the resting state of the receptor; the agonist-bound complex corresponds to the active state. Both subunits adopt an open conformation at rest, and only GBR1 closes on agonist-induced receptor activation. The agonists and antagonists are anchored in the interdomain crevice of GBR1 by an overlapping set of residues. An antagonist confines GBR1 to the open conformation of the inactive state, whereas an agonist induces its domain closure for activation. Our data reveal a unique activation mechanism for GABA_B receptor that involves the formation of a novel heterodimer interface between subunits.

GABA is the predominant inhibitory neurotransmitter in the central nervous system. Metabotropic GABA_B receptor is a G-protein-coupled receptor that mediates slow and prolonged synaptic inhibition through G_i or G_o protein^{1,2}. Presynaptic GABA_B receptor suppresses neurotransmitter release, and postsynaptic GABA_B receptor causes hyperpolarization of neurons^{1,2}. Malfunction of GABA_B receptor can lead to various neurological disorders, including spasticity and epilepsy^{1–3}. Baclofen, a selective GABA_B agonist, is used clinically to treat muscle spasticity associated with multiple sclerosis, cerebral palsy and spinal cord injury^{1–3}.

GABA_B receptor belongs to the distinct class-C family of G-protein-coupled receptors⁴. Ligand binding to these receptors takes place within a large extracellular Venus flytrap (VFT) module that has sequence homology to bacterial periplasmic amino-acid-binding proteins⁴. Unlike metabotropic glutamate receptors (mGluRs) and extracellular-calcium-sensing receptor, which function as disulphide-tethered homodimers^{5–8}, GABA_B and taste receptors act as heterodimers^{9–16}.

GABA_B receptor functions as a heterodimeric assembly of the subunits GBR1 and GBR2 (refs 9–12, 14). GBR2 facilitates cell surface expression of GBR1 by masking an endoplasmic reticulum retention signal of GBR1 (refs 17, 18). GBR1 is responsible for ligand recognition through its extracellular domain^{19,20}. Although GBR2 does not bind any known GABA_B ligand^{9–11,21}, its ectodomain directly interacts with the GBR1 ectodomain to enhance agonist affinity^{10,11,22–26} and is required for receptor activation^{22,25,27}. Finally, the transmembrane domain of GBR2 is responsible for G-protein coupling^{22,25,28–32}.

Most of the current knowledge about the structures of class-C G-protein-coupled receptors derives from homodimeric mGluRs. The ectodomain structures of three mGluR subtypes have been determined with and without ligand^{33–35}. Here we assembled a stable heterodimeric complex of the human GBR1 and GBR2 ectodomains, and determined its crystal structure in the absence of ligand and in the presence of various agonists and antagonists. Together with our mutational data, these structures provide insights into the molecular mechanisms of receptor heterodimerization, ligand recognition and receptor activation.

Structures of GABA_B heterodimer

The extracellular VFT module of human GBR1b (GBR1b_{VFT}) and GBR2 (GBR2_{VFT}) were co-secreted as a heterodimeric complex from

insect cells (Supplementary Fig. 1). The GBR1b_{VFT}:GBR2_{VFT} heterodimer binds various agonists and antagonists with the same rank order of affinities as the full-length receptor, indicating that it is physiologically relevant²⁶.

We determined the crystal structure of the GBR1b_{VFT}:GBR2_{VFT} complex in the apo form, when bound to six different antagonists (CGP54626_{ANT}, CGP46381_{ANT}, CGP35348_{ANT}, SCH50911_{ANT}, (S)-2-OH-saclofen_{ANT} and (R)-phaclofen_{ANT}) and when bound to two different agonists (endogenous ligand GABA and clinical drug (R)-baclofen_{AGO}) (Supplementary Table 1). Each structure consists of a non-covalent heterodimer of GBR1b_{VFT} and GBR2_{VFT}, in which the protomers are bound to each other such that they are side by side and facing opposite directions (Fig. 1 and Supplementary Fig. 2). All of the agonists and antagonists bind in the crevice between the LB1 and LB2 domains of GBR1b_{VFT}.

GBR1b_{VFT} and GBR2_{VFT} have similar overall structures, in agreement with their sequence homology (33% identity) (Supplementary Fig. 3). Both subunits have a bi-lobed architecture related to that found in mGluRs^{33–35}, natriuretic peptide receptors^{36,37}, ionotropic glutamate receptors^{38–40} and periplasmic amino-acid-binding proteins⁴¹. However, the extracellular domains of GBR1b and GBR2 lack the cysteine-rich region found at the carboxy-terminal end of mGluR ectodomains. Each GABA_B subunit contains two distinct domains, LB1 and LB2. The individual LB1 and LB2 domains of the two subunits are highly correlated with each other.

Despite similarities, the GBR1b_{VFT} and GBR2_{VFT} structures have different interdomain arrangements, consistent with their disparate ligand-binding characteristics (Fig. 2a, b). The ligand-binding subunit, GBR1b_{VFT}, can oscillate between open and closed conformations, in which the more compact closed conformation is associated with agonist binding. In contrast, the non-ligand-binding subunit, GBR2_{VFT}, has nearly identical conformations with and without dimer partner GBR1_{VFT}.

In the crystal structure of apo-GBR1b_{VFT}:GBR2_{VFT}, both subunits adopt an open conformation when compared with the known structures of mGluRs^{33–35} (Supplementary Fig. 4). All six antagonist-bound structures closely resemble that of the apo complex, both in the arrangement of the heterodimer and in the structures of the individual subunits

¹Department of Pharmacology, Columbia University, New York, New York 10032, USA. ²Department of Pathology & Cell Biology, Columbia University, New York, New York 10032, USA.

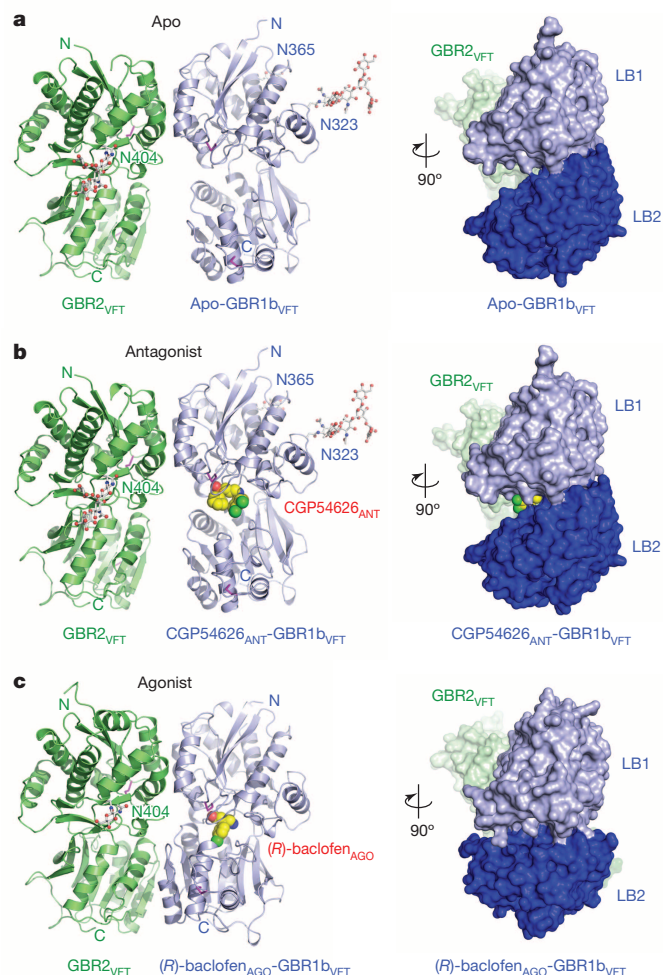


Figure 1 | Crystal structures of the GBR1b_{VFT}:GBR2_{VFT} complex. **a**, Apo structure; **b**, CGP54626_{ANT}-bound structure; **c**, (*R*)-baclofen_{AGO}-bound structure. Each complex is shown in two views related by a 90° rotation about the vertical axis. Front view (left panel) is shown as a ribbon diagram; side view (right panel) is presented as a molecular surface. GBR1b_{VFT} and GBR2_{VFT} are coloured blue and green, respectively. The observed carbohydrates are shown as ball-and-stick models in grey. Disulphide bridges are in magenta. The ligands are displayed as space-filling models.

(Supplementary Fig. 4). The ligand-binding cleft of GBR1b_{VFT} stays open with each bound antagonist. In addition, GBR2_{VFT} remains wide open with an empty interdomain cleft. This open–open configuration of the apo and antagonist-bound structures corresponds to the resting (or inactive) state of the heterodimeric receptor.

Agonist binding causes large conformational changes within the heterodimeric complex. First, both the agonists GABA and (*R*)-baclofen_{AGO} induce domain closure of GBR1b_{VFT}, as previously predicted⁴² (Fig. 2a). When the LB1 domains of apo and agonist-bound GBR1b_{VFT} are superimposed, their LB2 domains can be related by a 29° rotation about a nearly horizontal interdomain axis. Because the rotational axis has a slight vertical offset, this transformation also brings the LB2 domain of GBR1b_{VFT} into close contact with the LB2 domain of GBR2_{VFT} to form a large heterodimer interface unique to the active state.

Second, GBR2_{VFT} remains open in the agonist-bound state, consistent with our previous prediction that GBR2_{VFT} has a constitutively open conformation²⁶. Nevertheless, the LB2 domain of GBR2_{VFT} undergoes a twist motion of 9° around a nearly vertical axis, and moves towards the LB2 domain of GBR1b_{VFT} to form new heterodimeric contacts (Fig. 2b).

Finally, the substantial rearrangement of the LB2 domains from the apo to the agonist-bound state shortens the distance between the C termini of the two subunits from 45 to 32 Å (Fig. 2c, d and Supplementary

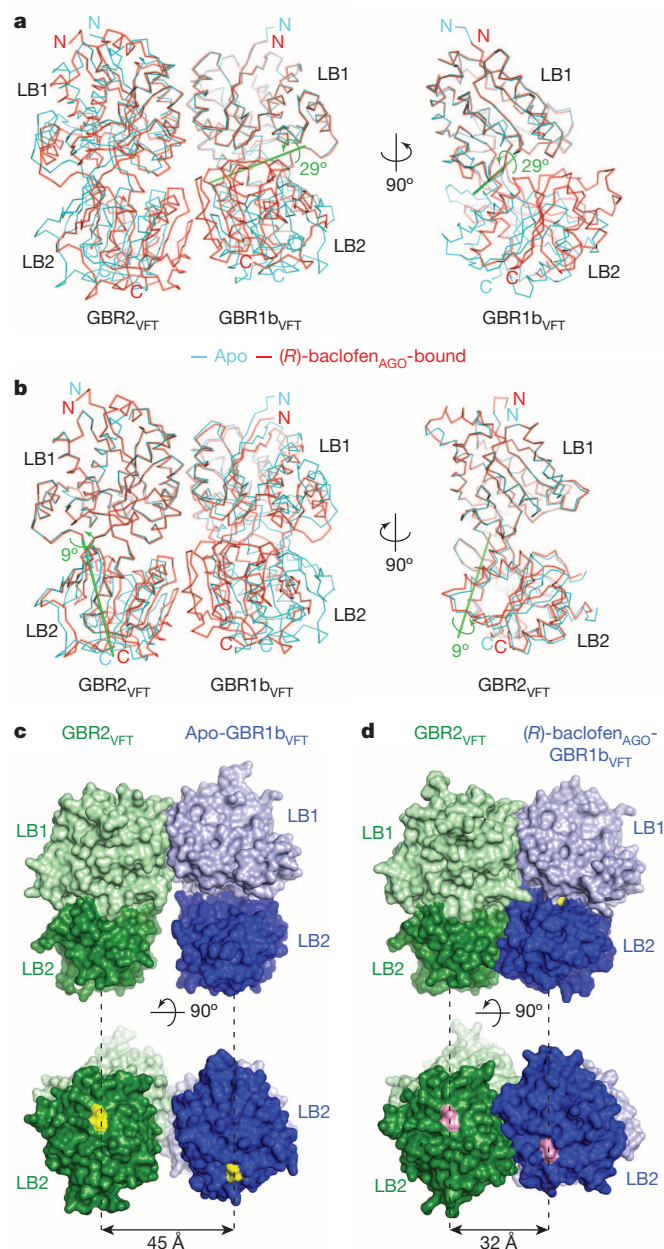


Figure 2 | Agonist-induced conformational changes. **a**, **b**, Superposition of apo (cyan) and (*R*)-baclofen_{AGO}-bound (red) complexes based on the LB1 domain of GBR1b_{VFT} (**a**) or GBR2_{VFT} (**b**). Side view is shown on the right for the superimposed GBR1b_{VFT} (**a**) and GBR2_{VFT} (**b**) subunits. Green line is the axis of rotation that relates the LB2 domains of apo and agonist-bound GBR1b_{VFT} (rotation, $\chi = 29^\circ$; screw translation, $\tau_\chi = 0.2 \text{ \AA}$) (**a**) and apo and agonist-bound GBR2_{VFT} ($\chi = 9^\circ$, $\tau_\chi = 0.1 \text{ \AA}$) (**b**). **c**, **d**, Surface representation of apo (**c**) and (*R*)-baclofen_{AGO}-bound (**d**) GBR1b_{VFT}:GBR2_{VFT} in front view (top) and bottom view (bottom). Distances between C termini of the two subunits (yellow in apo structure; pink in (*R*)-baclofen_{AGO}-bound structure) are marked by dashed lines.

Fig. 5). This decrease in the separation between membrane-proximal LB2 domains may lead to changes in the relative orientation of the transmembrane domains. In summary, both agonist-bound GBR1b_{VFT}:GBR2_{VFT} complexes adopt a closed–open structural arrangement, which corresponds to the active state of the receptor (Supplementary Fig. 4).

Common subunit–subunit interactions

In both the resting and active states, GBR1b_{VFT} and GBR2_{VFT} interact through their LB1 domains (Supplementary Figs 6 and 7). In the apo

and antagonist-bound structures, the subunit association is exclusively facilitated by this LB1–LB1 contact. The heterodimer buries more than 1,400 Å² of solvent-accessible surface area and has exceptionally high interfacial shape correlation (Supplementary Table 2).

The LB1–LB1 interaction is mediated by the B and C helices of both subunits (Fig. 3a). The heterodimer interface can be divided into three regions (Fig. 3b). Site I is located at the centre of the interface, and it is flanked, one on either side, by sites II and III.

Site I consists of a central hydrophobic patch surrounded by hydrogen bonds. The heterodimer contacts within this site are highly conserved in all of the GBR1b_{VFT}:GBR2_{VFT} structures. In particular, it features three deeply buried tyrosine residues (Tyr 113 and Tyr 117 of GBR1b_{VFT} and Tyr 118 of GBR2_{VFT}) that are critical for heterodimer interaction and receptor activation²⁶. These tyrosine residues participate in aromatic stacking interactions, and form interfacial hydrogen bonds. Together with the adjacent lysine and tryptophan residues, they are responsible for the majority of hydrophobic contacts at the LB1–LB1 heterodimer interface.

Interactions at site II are mostly hydrogen bonds, and include a universal salt bridge (GBR1b_{VFT}-R141:GBR2_{VFT}-D109) as well as a conserved hydrogen bond (GBR1b_{VFT}-E138:GBR2_{VFT}-N110). Site III consists predominantly of water-mediated contacts, and is the most variable part of the LB1–LB1 interface.

Agonist-induced heterodimer interface

Agonist binding induces the formation of an additional heterodimer interface between the LB2 domains of the GBR1b_{VFT} and GBR2_{VFT} subunits (Supplementary Fig. 7). This is consistent with our calorimetry measurements showing that GBR2_{VFT} has higher affinity for agonist-bound than antagonist-bound GBR1b_{VFT} (ref. 26). The LB2–LB2 interface buries more than 1,300 Å² of solvent-accessible surface area, has poor shape complementarity and is dominated by polar interactions (Supplementary Table 2).

The LB2–LB2 interaction is mediated by two strand–loop–helix motifs from each LB2 domain (Fig. 3c). Neighbouring strands f and g are part of the central β-sheet in LB2, and helices F and G flank the β-sheet. The heterodimer contacts consist primarily of hydrogen bonds, some of which are mediated by water molecules. The interface can be divided into three adjacent areas (Fig. 3d). Sites IV and V each features a large cluster of hydrogen bonds, and site VI mostly consists of isolated contacts. The GBR2_{VFT} residue Asn 213 is located at the intersection of sites IV and V, and it bridges the hydrogen-bond networks within these

two regions. In addition, a minor LB2–LB1 contact involving helix D of GBR2_{VFT} is formed at the edge of site IV.

To confirm the importance of the LB2–LB2 heterodimer interface to receptor activation, we carried out alanine-scanning mutagenesis of the interfacial residues. We identified several polar residues from each subunit that are critical to agonist-dependent G_i-protein activity (Supplementary Fig. 7). These include the GBR1b_{VFT} residues Thr 198, Glu 201 and Ser 225, and the GBR2_{VFT} residues Asp 204, Gln 206, Asn 213 and Ser 233. All of these residues are engaged in multiple interfacial hydrogen bonds at the LB2–LB2 interface. This reliance on hydrophilic interactions to form a distinct subunit interface in the active state allows the receptor to dissociate readily on returning to its resting state. Previous studies have also shown that introduction of a large N-linked glycan into the LB2 domain of either GABA_B subunit inhibits agonist-induced receptor activation⁴³.

Ligand recognition

All of the antagonists are derivatives of GABA, and have the general structure of a γ-amino acid. The receptor–antagonist interactions are mediated largely by hydrogen bonds (Fig. 4a, b and Supplementary Fig. 8). First, each antagonist is anchored at the crevice of GBR1b_{VFT} by two sets of hydrogen bonds. The α-acid group at one end forms hydrogen bonds with the LB1 residues Ser 130 and Ser 153, and the γ-amino group at the other end is hydrogen-bonded to His 170 and Glu 349. Second, Trp 65 makes van der Waals contacts with all of the antagonists. Third, the β-hydroxyl substituent of CGP54626_{ANT} and (S)-2-OH-saclofen_{ANT} makes additional hydrogen bonds with the receptor that are specific to these antagonists. Finally, all of the antagonists except SCH50911_{ANT} and (R)-phaclofen_{ANT} participate in water-mediated interaction with Ser 131. These extensive contacts indicate that the LB1 domain is primarily responsible for anchoring antagonist.

In contrast, the interaction between the LB2 domain and bound antagonist is sparse and varies among the different antagonists (Supplementary Fig. 8). Only two antagonists, CGP54626_{ANT} and SCH50911_{ANT}, directly contact Trp 278 of LB2 through a large γ-substituent. As a result of this additional LB2 interaction, both compounds have higher binding affinity to GABA_B receptor than the other antagonists reported here³. This suggests that the LB2 domain has an auxiliary role in antagonist recognition, and enhances the potency of selective antagonists.

GABA_B receptor recognizes both the agonists GABA and (R)-baclofen_{AGO} in essentially the same manner (Fig. 4c, d and Supplementary Fig. 9). (R)-baclofen_{AGO} is a derivative of GABA, and contains a

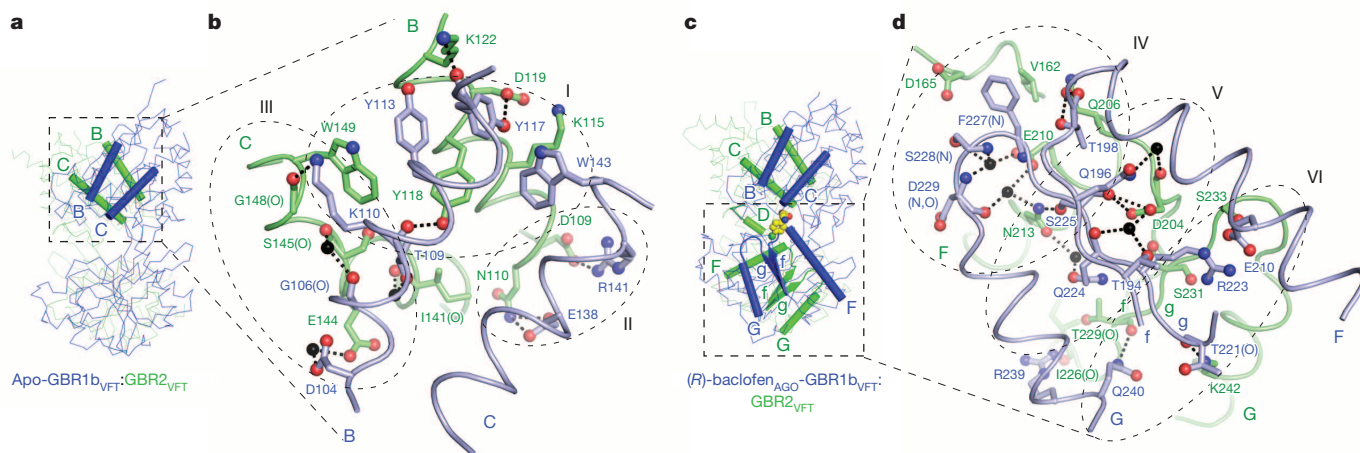


Figure 3 | Heterodimer interface. **a**, Structure of apo-GBR1b_{VFT}:GBR2_{VFT} with the elements involved in heterodimer formation highlighted by ribbons (LB1–LB1: B and C helices). **b**, Specific contacts at the LB1–LB1 heterodimer interface of apo-GBR1b_{VFT}:GBR2_{VFT}. The interface area is divided into three regions, I, II and III (thin dashed lines). Thick dashed lines indicate hydrogen bonds. **c**, Structure of (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} showing the

elements involved in heterodimer formation (LB1–LB1: B and C helices; LB2–LB2: F and G helices, f and g strands, and connecting loops). **d**, Specific contacts at the LB2–LB2 heterodimer interface of (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT}. The interface area is divided into three regions, IV, V and VI (thin dashed lines). Thick dashed lines indicate hydrogen bonds.

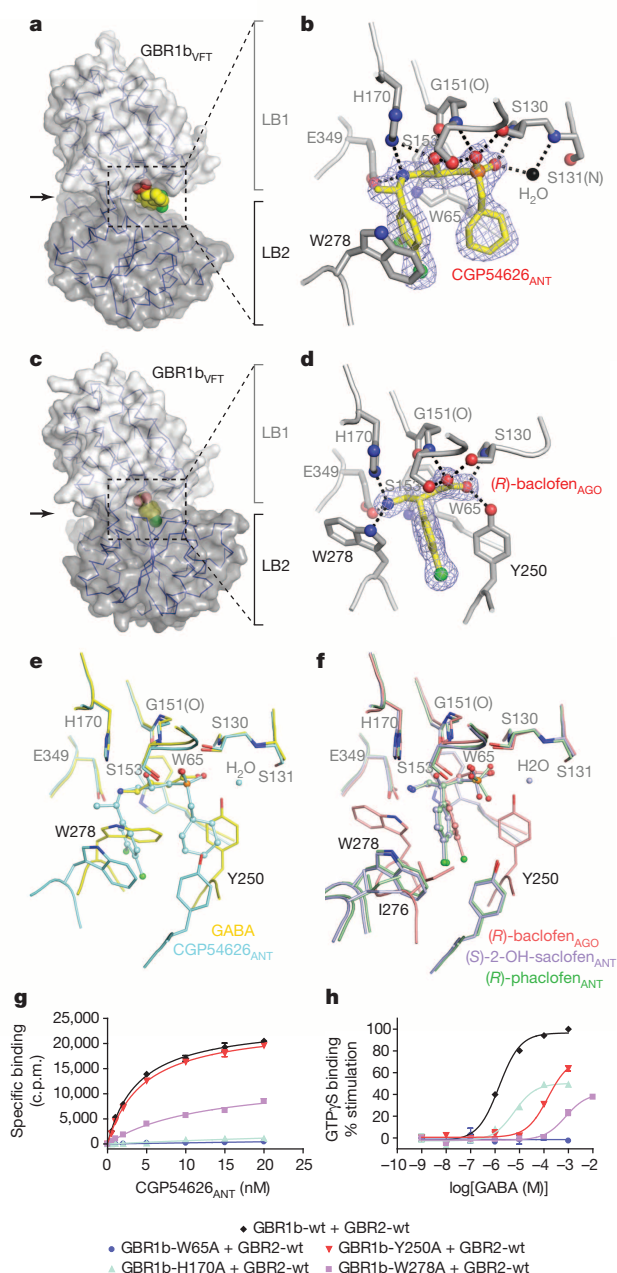


Figure 4 | Ligand recognition by GBR1b_{VFT}. **a, c**, Molecular surface of GBR1b_{VFT} bound to antagonist CGP54626_{ANT} (**a**) or agonist (R)-baclofen_{AGO} (**c**). Ligand is displayed as a space-filling model. **b, d**, Specific contacts between GBR1b_{VFT} (grey) and CGP54626_{ANT} (yellow) (**b**) or (R)-baclofen_{AGO} (**d**), viewed in the direction of the arrow in **a** (**b**) or **c** (**d**). Mesh represents the final $2F_o - F_c$ electron density map contoured at 1σ . Hydrogen bonds are represented by thick dashed lines. **e**, Comparison of the binding sites of agonist (R)-baclofen_{AGO} and two related antagonists, (S)-2-OH-saclofen_{ANT} and (R)-phaclofen_{ANT}. **f**, Comparison of the binding sites of agonist (R)-baclofen_{AGO} and two related antagonists, (S)-2-OH-saclofen_{ANT} and (R)-phaclofen_{ANT}. **g, h**, Dose-dependent [³H]CGP54626_{ANT} binding (**g**) and GABA-stimulated [³⁵S]GTPγS binding (**h**) in membranes from cells expressing wild-type GABA_B receptor (GBR1b-wt + GBR2-wt) or the combination of GBR2-wt and various GBR1b mutants. c.p.m., counts per minute. Data points represent average \pm s.e.m. of triplicate measurements.

chlorophenyl substituent at the β -position. Like the antagonists, each agonist is secured by two hydrogen-bond networks, one at each end of the molecule. Furthermore, a common set of LB1 residues is involved in binding the two ends of all of the agonists and antagonists. Unlike the antagonists, both agonists also directly contact two key residues of the LB2 domain, Tyr 250 and Trp 278. In addition, the two tryptophan

residues Trp 65 and Trp 278 make extensive van der Waals contacts with both GABA and (R)-baclofen_{AGO}. Therefore, both the LB1 and LB2 domains are required for agonist recognition.

The binding sites of GABA and (R)-baclofen_{AGO} differ in the side-chain conformation of the LB2 residue Trp 278 (Supplementary Fig. 9). Relative to its orientation in the GABA-bound complex, the indole ring of Trp 278 is flipped by $\sim 170^\circ$ to accommodate the β -chlorophenyl substituent of (R)-baclofen_{AGO}, which forms aromatic-ring-stacking interactions with both Tyr 250 and Trp 278. In contrast, GABA makes van der Waals contact with Trp 278 alone through its aliphatic backbone. The conformational adaptability of Trp 278 provides a mechanism by which the receptor recognizes structurally different ligands while maintaining ligand-binding specificity and affinity.

Agonist action versus antagonist action

The function of a GABA_B agonist is to stabilize the closed conformation of GBR1b_{VFT}, whereas that of an antagonist is to confine the GBR1b_{VFT} subunit to the open configuration (Supplementary Fig. 10). Agonist-bound GBR1b_{VFT} has a closed cleft; the agonist is buried and inaccessible to the bulk solvent. In contrast, antagonist-bound GBR1b_{VFT} has an open cleft, and the antagonist is solvent accessible.

The presence of a bulky substituent in each antagonist inhibits domain closure of GBR1b_{VFT}. The highly potent antagonist CGP54626_{ANT} contains an α -cyclohexyl group and a γ -dichlorophenyl group. The adverse interactions of these moieties with Tyr 250 and Trp 278 would be expected to prevent the LB1 and LB2 domains from approaching each other (Fig. 4e). Similarly, each of the other antagonists CGP46381_{ANT}, CGP35348_{ANT} and SCH50911_{ANT} has a bulky substituent at either the α -position or the γ -position to block GBR1b_{VFT} domain closure (Supplementary Figs 8 and 9). Although the antagonists (S)-2-OH-saclofen_{ANT} and (R)-phaclofen_{ANT} are structurally analogous to the agonist (R)-baclofen_{AGO}, their α -acid motifs assume a tetrahedral coordination geometry that is incompatible with the active-state conformation of Tyr 250 (Fig. 4f). Furthermore, the α -substituents push the β -chlorophenyl ring towards the γ -amino end of each antagonist, thereby generating potential steric interactions with Ile 276 and Trp 278 to prevent GBR1b_{VFT} domain closure.

All of the residues at the ligand-binding site are conserved within GBR1 sequences across different species (Supplementary Fig. 11). Some of the ligand-binding residues, including Ser 130, Gly 151, Ser 153 and Glu 349 of GBR1b, have been implicated by previous studies^{21,26,44–46}.

The LB1 residues are required for both agonist and antagonist recognition. We found that the Trp65Ala substitution caused substantial loss of ligand binding and receptor function (Fig. 4g, h). The His170Ala mutation essentially abolished antagonist binding, and lowered the maximum agonist-induced [³⁵S]GTPγS binding to half that of wild-type level (Fig. 4g, h). These data indicate that both Trp 65 and His 170 are indispensable for ligand recognition.

The LB2 residues are essential for agonist binding. First, the Trp278Ala mutant retained the ability to bind the antagonist [³H]CGP54626_{ANT}, although with decreased potency (Fig. 4g). This is consistent with the auxiliary role of Trp 278 in antagonist recognition. However, this mutation is detrimental to receptor activation, because it not only reduced the maximum GABA-dependent [³⁵S]GTPγS binding, but also increased the half-maximum effective concentration (EC_{50}) of GABA by more than 500-fold (Fig. 4h). Second, the Tyr250Ala mutation had no effect on antagonist binding, in agreement with our structural observations (Fig. 4g). However, it decreased the agonist response, and increased the EC_{50} of GABA by more than 100-fold (Fig. 4h). These data indicate that both Tyr 250 and Trp 278 are critical to agonist recognition.

Implications for receptor activation

Structural comparison indicates that the concept of major intersubunit relocation that holds for the activation of mGluRs cannot be applied to GABA_B receptor. The extracellular domains of these receptors share a common mode of dimerization through their LB1 domains (Supplementary

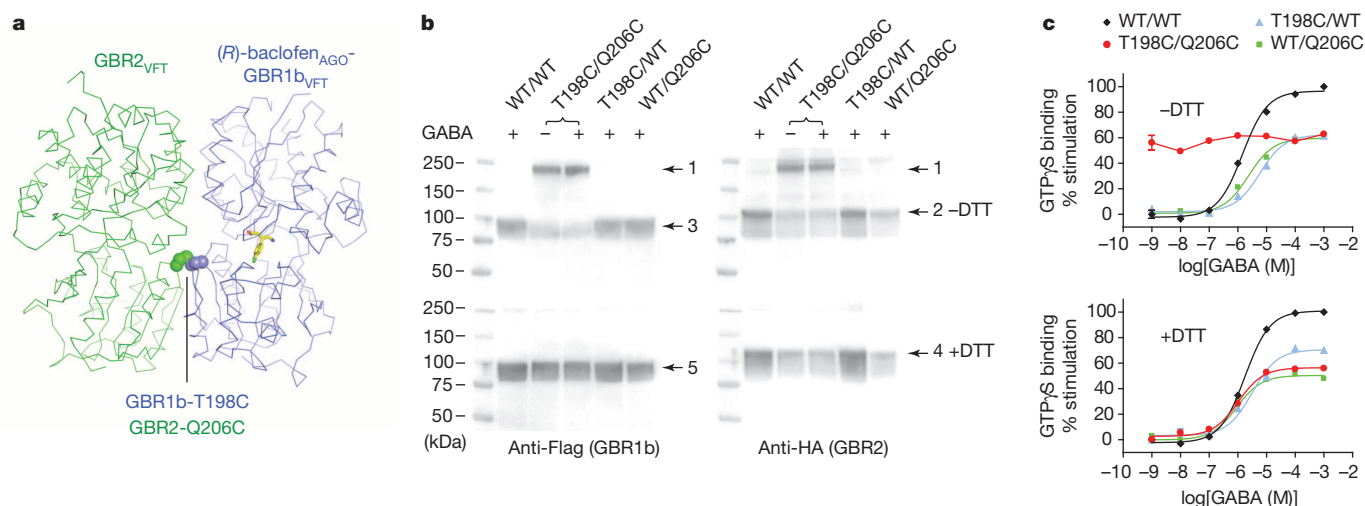


Figure 5 | Constitutive activity of disulphide-tethered GBR1b:GBR2 heterodimer. **a**, Position of cysteine mutations (spheres) at the LB2–LB2 heterodimer interface of (R)-baclofen_{AGO}–GBR1b_{VFT}:GBR2_{VFT}. **b**, Western-blot analysis of membranes from cells expressing different combinations of wild-type (WT) and mutant GABA_B receptor subunits (GBR1b–Thr198Cys, abbreviated as T198C; GBR2–Gln206Cys, abbreviated as Q206C). The samples were assayed in the presence of 10 mM GABA under reducing (+DTT) and non-reducing (–DTT) conditions. The double cysteine

Figs 12 and 13). The resting and active configurations of mGluRs differ by a 70° rotation in dimer orientation^{33–35}. Both closed–open and closed–closed conformations have been reported for activated mGluRs^{33–35}, although full activation requires the closure of both protomers⁴⁷. In contrast, the heterodimeric LB1–LB1 interface of GABA_B receptor undergoes a minor 5° rotation in dimer orientation on agonist binding, and the receptor adopts only a closed–open active conformation. Our data indicate that activation of GABA_B receptor involves the formation of a novel LB2–LB2 heterodimer interface.

We carried out disulphide crosslinking studies⁴⁸ to determine the physiological relevance of the LB2–LB2 interaction in full-length receptor. On the basis of the active-state structure of GBR1b_{VFT}:GBR2_{VFT}, we introduced cysteine mutations into a residue pair across the LB2–LB2 dimer interface (GBR1–Thr198Cys and GBR2–Gln206Cys), which had the proximity and geometry required for disulphide formation (Fig. 5a). Western-blot analysis indicates that co-expression of wild-type GBR1b and GBR2 or the combination of a single cysteine mutant with its wild-type partner in mammalian cells produced monomeric protein bands in the presence of GABA under both reducing and non-reducing conditions (~95 kDa for GBR1b; ~115 kDa for GBR2) (Fig. 5b). In contrast, co-expression of the cysteine mutant pair yielded a heterodimeric protein band (~210 kDa) under non-reducing conditions (Fig. 5b). This band was recognized by both anti-Flag and anti-HA antibodies, which were used to detect differentially tagged GBR1b and GBR2 subunits. Furthermore, it was observed in the absence of ligand and in the presence of the agonist GABA. These observations indicate the spontaneous formation of a disulphide-tethered GBR1b:GBR2 heterodimer, and confirm that the LB2–LB2 interface observed in the active-state GBR1b_{VFT}:GBR2_{VFT} structure is also present in free and agonist-bound native GABA_B receptor.

To determine the functional effects of locking the LB2–LB2 interface, we measured agonist-dependent G_i-protein activation of different combinations of wild-type and cysteine mutant receptors (Fig. 5c). For the wild-type receptor and single cysteine mutants, application of GABA led to stimulation of [³⁵S]GTPγS binding both in the absence and presence of dithiothreitol (DTT). In contrast, the double cysteine mutant exhibited constitutive activity under non-reducing conditions, and addition of GABA did not further increase its functional activity (Fig. 5c). This indicates that the intersubunit disulphide bond holds the receptor in a fully active form. Indeed, on reduction of the disulphide

mutant (Thr198Cys Gln206Cys) was also analysed in the absence of ligand. GBR1b and GBR2 were detected by anti-Flag and anti-HA antibodies, respectively. Arrow 1, GBR1b–GBR2 heterodimer; arrows 2 and 4, GBR2 monomer; arrows 3 and 5, GBR1b monomer. **c**, GABA-stimulated, dose-dependent [³⁵S]GTPγS binding in membranes from cells expressing wild-type or various cysteine mutant receptors in the presence or absence of DTT. Data points represent average ± s.e.m. of triplicate measurements.

bond, the double cysteine mutant receptor lost its constitutive activity, but regained sensitivity to GABA to a level comparable to that of a single cysteine mutant (Fig. 5c). Our data demonstrate that formation of the LB2–LB2 interface is both necessary and sufficient for GABA_B receptor activation.

In the conformational equilibrium of GABA_B receptor, an antagonist maintains the inactive conformation of the receptor, whereas an agonist stabilizes its active conformation (Supplementary Fig. 14). Agonist binding to GABA_B receptor induces domain closure in the GBR1 subunit, an expansion of the heterodimer interaction to include a large LB2–LB2 interface, and a decrease in the separation between the membrane-proximal LB2 domains. Because receptor function is not affected by alterations in the peptide linker between the VFT and transmembrane domains of each subunit²⁷, these changes would probably be directly relayed to the transmembrane domains. We expect that the transmembrane domains of the GABA_B subunits exist as pre-formed heterodimers on the cell surface because both the extracellular and the intracellular components form stable heterodimers^{12,23,24,26,49}. Therefore, agonist-induced conformational changes may lead to a rearrangement of the transmembrane-domain heterodimer for signal transduction across the membrane. This novel activation mechanism would be, as of yet, unique to inhibitory GABA_B receptor.

METHODS SUMMARY

The extracellular VFT module of human GBR1b and GBR2 were co-secreted from baculovirus-infected insect cells, and purified by anti-Flag M2 antibody affinity chromatography followed by size-exclusion chromatography. Apo and antagonist-bound GBR1b_{VFT}:GBR2_{VFT} complexes were crystallized in the *P*₂₁ space group. Agonist-bound complexes were crystallized in the *P*₂₁₂₁2 space group. All of the structures were solved by molecular replacement. Radioligand binding and agonist-stimulated [³⁵S]GTPγS binding were measured using HEK293 cells co-transfected with complementary DNAs encoding full-length GBR1b and GBR2.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 January; accepted 30 September 2013.

Published online 4 December 2013.

1. Bettler, B., Kaupmann, K., Mosbacher, J. & Gassmann, M. Molecular structure and physiological functions of GABA(B) receptors. *Physiol. Rev.* **84**, 835–867 (2004).

2. Bowery, N. G. *et al.* International Union of Pharmacology. XXXIII. Mammalian gamma-aminobutyric acid(B) receptors: structure and function. *Pharmacol. Rev.* **54**, 247–264 (2002).
3. Froestl, W. Chemistry and pharmacology of GABA_B receptor ligands. *Adv. Pharmacol.* **58**, 19–62 (2010).
4. Pin, J. P. *et al.* The activation mechanism of class-C G-protein coupled receptors. *Biol. Cell* **96**, 335–342 (2004).
5. Romano, C., Yang, W. L. & O'Malley, K. L. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J. Biol. Chem.* **271**, 28612–28616 (1996).
6. Okamoto, T. *et al.* Expression and purification of the extracellular ligand binding region of metabotropic glutamate receptor subtype 1. *J. Biol. Chem.* **273**, 13089–13096 (1998).
7. Tsuji, Y. *et al.* Cryptic dimer interface and domain organization of the extracellular region of metabotropic glutamate receptor subtype 1. *J. Biol. Chem.* **275**, 28144–28151 (2000).
8. Bai, M., Trivedi, S. & Brown, E. M. Dimerization of the extracellular calcium-sensing receptor (CaR) on the cell surface of CaR-transfected HEK293 cells. *J. Biol. Chem.* **273**, 23605–23610 (1998).
9. Jones, K. A. *et al.* GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* **396**, 674–679 (1998).
10. Kaupmann, K. *et al.* GABA(B)-receptor subtypes assemble into functional heteromeric complexes. *Nature* **396**, 683–687 (1998).
11. White, J. H. *et al.* Heterodimerization is required for the formation of a functional GABA(B) receptor. *Nature* **396**, 679–682 (1998).
12. Kuner, R. *et al.* Role of heteromer formation in GABA_B receptor function. *Science* **283**, 74–77 (1999).
13. Milligan, G. G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. *Br. J. Pharmacol.* **158**, 5–14 (2009).
14. Ng, G. Y. *et al.* Identification of a GABA_B receptor subunit, gb2, required for functional GABA_B receptor activity. *J. Biol. Chem.* **274**, 7607–7610 (1999).
15. Nelson, G. *et al.* Mammalian sweet taste receptors. *Cell* **106**, 381–390 (2001).
16. Nelson, G. *et al.* An amino-acid taste receptor. *Nature* **416**, 199–202 (2002).
17. Margeta-Mitrovic, M., Jan, Y. N. & Jan, L. Y. A trafficking checkpoint controls GABA(B) receptor heterodimerization. *Neuron* **27**, 97–106 (2000).
18. Pagano, A. *et al.* C-terminal interaction is essential for surface trafficking but not for heteromeric assembly of GABA(b) receptors. *J. Neurosci.* **21**, 1189–1202 (2001).
19. Kaupmann, K. *et al.* Expression cloning of GABA(B) receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
20. Malitschek, B. *et al.* The N-terminal domain of gamma-aminobutyric Acid(B) receptors is sufficient to specify agonist and antagonist binding. *Mol. Pharmacol.* **56**, 448–454 (1999).
21. Kniazeff, J., Galvez, T., Labesse, G. & Pin, J. P. No ligand binding in the GB2 subunit of the GABA(B) receptor is required for activation and allosteric interaction between the subunits. *J. Neurosci.* **22**, 7352–7361 (2002).
22. Galvez, T. *et al.* Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA(B) receptor function. *EMBO J.* **20**, 2152–2159 (2001).
23. Liu, J. *et al.* Molecular determinants involved in the allosteric control of agonist affinity in the GABA_B receptor by the GABA_B2 subunit. *J. Biol. Chem.* **279**, 15824–15830 (2004).
24. Nomura, R., Suzuki, Y., Kakizuka, A. & Jingami, H. Direct detection of the interaction between recombinant soluble extracellular regions in the heterodimeric metabotropic gamma-aminobutyric acid receptor. *J. Biol. Chem.* **283**, 4665–4673 (2008).
25. Monnier, C. *et al.* Trans-activation between 7TM domains: implication in heterodimeric GABA(B) receptor activation. *EMBO J.* **30**, 32–42 (2011).
26. Geng, Y. *et al.* Structure and functional interaction of the extracellular domain of human GABA_B receptor GBR2. *Nature Neurosci.* **15**, 970–978 (2012).
27. Margeta-Mitrovic, M., Jan, Y. N. & Jan, L. Y. Ligand-induced signal transduction within heterodimeric GABA(B) receptor. *Proc. Natl Acad. Sci. USA* **98**, 14643–14648 (2001).
28. Margeta-Mitrovic, M., Jan, Y. N. & Jan, L. Y. Function of GB1 and GB2 subunits in G protein coupling of GABA(B) receptors. *Proc. Natl Acad. Sci. USA* **98**, 14649–14654 (2001).
29. Robbins, M. J. *et al.* GABA(B2) is essential for g-protein coupling of the GABA(B) receptor heterodimer. *J. Neurosci.* **21**, 8043–8052 (2001).
30. Duthey, B. *et al.* A single subunit (GB2) is required for G-protein activation by the heterodimeric GABA(B) receptor. *J. Biol. Chem.* **277**, 3236–3241 (2002).
31. Havlickova, M. *et al.* The intracellular loops of the GB2 subunit are crucial for G-protein coupling of the heteromeric gamma-aminobutyrate B receptor. *Mol. Pharmacol.* **62**, 343–350 (2002).
32. Pin, J. P. *et al.* Activation mechanism of the heterodimeric GABA(B) receptor. *Biochem. Pharmacol.* **68**, 1565–1572 (2004).
33. Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
34. Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H. & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd³⁺. *Proc. Natl Acad. Sci. USA* **99**, 2660–2665 (2002).
35. Muto, T., Tsuchiya, D., Morikawa, K. & Jingami, H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proc. Natl Acad. Sci. USA* **104**, 3759–3764 (2007).
36. van den Akker, F. *et al.* Structure of the dimerized hormone-binding domain of a guanylyl-cyclase-coupled receptor. *Nature* **406**, 101–104 (2000).
37. He, X., Chow, D., Martick, M. M. & Garcia, K. C. Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. *Science* **293**, 1657–1662 (2001).
38. Jin, R. *et al.* Crystal structure and association behaviour of the GluR2 amino-terminal domain. *EMBO J.* **28**, 1812–1823 (2009).
39. Karakas, E., Simorowski, N. & Furukawa, H. Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit. *EMBO J.* **28**, 3910–3920 (2009).
40. Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. The N-terminal domain of GluR6-subtype glutamate receptor ion channels. *Nature Struct. Mol. Biol.* **16**, 631–638 (2009).
41. Sack, J. S., Saper, M. A. & Quirocho, F. A. Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. *J. Mol. Biol.* **206**, 171–191 (1989).
42. Kniazeff, J. *et al.* Locking the dimeric GABA(B) G-protein-coupled receptor in its active state. *J. Neurosci.* **24**, 370–377 (2004).
43. Rondard, P. *et al.* Functioning of the dimeric GABA(B) receptor extracellular domain revealed by glycan wedge scanning. *EMBO J.* **27**, 1321–1332 (2008).
44. Galvez, T. *et al.* Mutagenesis and modeling of the GABA_B receptor extracellular domain support a venus flytrap mechanism for ligand binding. *J. Biol. Chem.* **274**, 13362–13369 (1999).
45. Galvez, T. *et al.* Mapping the agonist-binding site of GABA_B type 1 subunit sheds light on the activation process of GABA_B receptors. *J. Biol. Chem.* **275**, 41166–41174 (2000).
46. Galvez, T. *et al.* Ca²⁺ requirement for high-affinity gamma-aminobutyric acid (GABA) binding at GABA(B) receptors: involvement of serine 269 of the GABA(B)R1 subunit. *Mol. Pharmacol.* **57**, 419–426 (2000).
47. Kniazeff, J. *et al.* Closed state of both binding domains of homodimeric mGlu receptors is required for full activity. *Nature Struct. Mol. Biol.* **11**, 706–713 (2004).
48. Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).
49. Kammerer, R. A. *et al.* Heterodimerization of a functional GABA_B receptor is mediated by parallel coiled-coil alpha-helices. *Biochemistry* **38**, 13263–13269 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. A. Hendrickson and R. Kass for advice and support; I. Berger for the gift of pFBDM vector; K. Rajashankar, K. Perry, S. Banerjee, F. Murphy, I. Kourinov and D. Neau for help with data collection; Y. Chen for technical assistance; and M. Evelyn for reading the manuscript. This work was supported by the American Heart Association grant SDG0835183N and the National Institute of Health grant R01GM088454 (both to Q.R.F.). Q.R.F. is an Irma Hirsch Career Scientist, Pew Scholar, McKnight Scholar and Schaefer Scholar.

Author Contributions Q.R.F. initiated the study and designed the experiments; Y.G., Q.R.F., M.B., L.M. and F.W. performed experiments and analysed data; Q.R.F. and Y.G. wrote the paper.

Author Information Atomic coordinates and diffraction data for the structures reported here are deposited in the RCSB Protein Data Bank under accession codes 4MQE, 4MQF, 4MR7, 4MR8, 4MR9, 4MRM, 4MS1, 4MS3 and 4MS4. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Q.R.F. (qf13@columbia.edu).

METHODS

Protein expression and purification. The extracellular domains of human GBR1 and GBR2 were separately cloned into the pFBDM vector⁵⁰ for expression in baculovirus-infected insect cells. The GBR1 isoform GBR1b¹⁹ was used in this study. The GBR1b_{VFT} construct contained residues 48–459, with the signal peptide of baculovirus envelope surface glycoprotein gp67 attached at the N terminus and a Flag tag at the C terminus. The GBR2_{VFT} construct contained residues 1–466 and a C-terminal Flag tag, as previously described²⁶.

Sf9 insect cells were co-infected with recombinant GBR1b_{VFT} and GBR2_{VFT} baculoviruses at 23 °C for 96 h. The GBR1b_{VFT}:GBR2_{VFT} complex was purified from cell supernatant by anti-Flag antibody (M2) affinity chromatography followed by gel filtration chromatography (Superdex 200, GE Healthcare). The CGP54626_{ANT}-GBR1b_{VFT}:GBR2_{VFT} complex was produced in the presence of 10 µM CGP54626_{ANT} throughout expression and 20 µM CGP54626_{ANT} during purification. The (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex was expressed and purified in the presence of 100 µM (R)-baclofen, and the GABA-GBR1b_{VFT}:GBR2_{VFT} complex was produced in the presence of 100 µM GABA.

Crystallization and data collection. Crystals of the apo-GBR1b_{VFT}:GBR2_{VFT} complex were grown at 4 °C in 10% PEG 3350, 20% glycerol and 0.12 M Na acetate, pH 7.0. Crystals of various antagonist-bound GBR1b_{VFT}:GBR2_{VFT} complexes were obtained under the same condition as the apo complex. Specifically, the CGP54626_{ANT}-bound heterodimer was crystallized using protein that was purified in the presence of CGP54626_{ANT}. The apo-GBR1b_{VFT}:GBR2_{VFT} complex was also co-crystallized with 10 mM of each of the following antagonists: CGP46381_{ANT}, CGP35348_{ANT}, SCH50911_{ANT}, (R,S)-2-OH-saclofen_{ANT} and (R,S)-phaclofen_{ANT}. All of the crystals were directly frozen from drops.

The agonist-bound (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex was crystallized at 20 °C from 20% PEG 2000, 15% glycerol, 0.2 M NH₄Cl and 0.1 M Na cacodylate, pH 5.2, in the presence of 10 mM (R)-baclofen. Crystals of the GABA-GBR1b_{VFT}:GBR2_{VFT} complex were grown at 20 °C from 18% PEG 2000, 5% glycerol, 0.15 M NH₄Cl and 0.1 M Na cacodylate, pH 5.0, in the presence of 10 mM GABA. The crystals were frozen in a cryoprotecting solution containing 20% glycerol and all other components of the crystallization solution.

Native data for the different complexes were collected at the 24ID-C and 24ID-E beamlines of the Advanced Photon Source. Diffraction data for the apo complex and the CGP46381_{ANT}, CGP35348_{ANT}, SCH50911_{ANT} and GABA-bound complexes were integrated using XDS⁵¹ and scaled with SCALA⁵². Data for the CGP54626_{ANT}, (S)-2-OH-saclofen_{ANT}, (R)-phaclofen_{ANT} and (R)-baclofen_{AGO}-bound complexes were integrated and scaled using HKL2000⁵³.

Structure determination. The structure of the apo-GBR1b_{VFT}:GBR2_{VFT} complex was solved by molecular replacement. The position of GBR2_{VFT} was identified using the structure of free GBR2_{VFT} (PDB code 4F11; ref. ²⁶) as the search model. The location of GBR1b_{VFT} was found by using the individual LB1 and LB2 domains of GBR2_{VFT} as the search probes. A complete atomic model of the apo-GBR1b_{VFT}:GBR2_{VFT} complex was developed through a succession of manual building and iterative refinement. The final model contained the GBR1b_{VFT} residues 48–368 and 377–459, the GBR2_{VFT} residues 53–292, 300–379 and 385–466, and part of the Flag tag at the C termini of both subunits. Carbohydrate residues were also attached to Asn 323 and Asn 365 of GBR1b_{VFT}, and to Asn 404 of GBR2_{VFT}.

All of the antagonist-bound GBR1b_{VFT}:GBR2_{VFT} structures were solved by molecular replacement using the apo-GBR1b_{VFT}:GBR2_{VFT} structure as the search model. For each complex, the bound antagonist was modelled into the residual electron density map obtained in the final rounds of refinement. All of the antagonist-bound structures contained the GBR1b_{VFT} residues 48–368 and 377–459, the GBR2_{VFT} residues 53–292, 300–379 and 385–466, and part of the Flag tag at the C termini of both subunits. Carbohydrate residues were also attached to Asn 323 and Asn 365 of GBR1b_{VFT}, and to Asn 404 of GBR2_{VFT}. Although a racemic mixture of (R)-2-OH-saclofen_{ANT} and (S)-2-OH-saclofen_{ANT} was used for crystallization, only the (S)-2-OH-saclofen_{ANT} enantiomer was bound to GBR1b_{VFT} in the structure. Our observation is consistent with previous findings that the (S)-2-OH-saclofen_{ANT} enantiomer is the active antagonist⁵⁴. Similarly, we found that (R)-phaclofen was the active enantiomer, in agreement with previous studies⁵⁵.

The structure of the (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex was also determined by molecular replacement. The position of GBR2_{VFT} was found using the GBR2_{VFT} structure from the apo complex as the search model. The (R)-baclofen_{AGO}-bound GBR1b_{VFT} molecule was located using the individual LB1 and LB2 domains of apo-GBR1b_{VFT} as the search probes. A complete model of the (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex was constructed through iterative rounds of manual building and refinement. The GABA-bound GBR1b_{VFT}:GBR2_{VFT} structure was solved using the refined (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex structure as the search model. For each complex, the bound agonist was modelled into the residual electron density map obtained in the final rounds of refinement. The (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} complex contained

the GBR1b_{VFT} residues 50–368 and 377–459; the GABA-GBR1b_{VFT}:GBR2_{VFT} complex contained the GBR1b_{VFT} residues 50–84, 92–337, 344–368 and 377–459. Both agonist-bound structures contained the GBR2_{VFT} residues 50–291 and 302–466, and part of the Flag tag at the C termini of both GBR1b_{VFT} and GBR2_{VFT}. Carbohydrate residues were attached to Asn 404 of GBR2_{VFT}.

Molecular replacement searches were carried out using PHASER⁵⁶. Model building was performed with COOT⁵⁷. Structural refinement was executed using BUSTER⁵⁸. Ramachandran statistics were calculated for each structure using MOLPROBITY⁵⁹. Pairwise structural comparison was performed using LSQMAN⁶⁰. Software installation support was provided by SBGrid⁶¹.

Cell surface expression. Full-length human GBR1b and GBR2 were individually cloned into a pcDNA3.1(+) vector (Invitrogen) for expression in human embryonic kidney (HEK293) cells. A Flag tag was inserted after the signal peptide of GBR1b, and an HA tag was placed after the signal peptide of GBR2. Mutants of GBR1b and GBR2 were constructed using the QuikChange mutagenesis system (Stratagene).

HEK293 T/17 cells (ATCC) were co-transfected using Lipofectamine 2000 (Invitrogen) with the GBR1b and GBR2 plasmids. Cells permeabilized with 0.5% Triton X100 were used to determine the total expression levels of GBR1b and GBR2 in transfected cells. Untreated cells were used to determine the cell surface expression level of each subunit. The amount of surface protein detected for each construct was normalized to that found in the total cell lysate.

The cells were blocked with 5% milk, and then incubated with mouse anti-Flag M1 antibody (Sigma) as the primary antibody to measure GBR1b expression. Similarly, mouse anti-HA antibody HA.11 clone 16B12 (Covance) was used to detect GBR2. Donkey anti-mouse IRDye 800-labelled antibody (LiCor) was used as the secondary antibody in both cases. Fluorescent signals were measured with an Odyssey Infrared Imager (LiCor). The results of three independent experiments were used for statistical analysis. All of the mutants reported here were expressed on the cell surface at levels comparable to that of wild-type GABA_B receptor.

Agonist-stimulated [³⁵S]GTPγS binding. HEK293 T/17 cells were transiently transfected with full-length GBR1b and GBR2 plasmids. The cells were collected in 50 mM HEPES, pH 7.4, to obtain the membrane fraction. Membranes were suspended in an assay buffer containing 50 mM Tris, pH 7.7, 100 mM NaCl, 12 mM MgCl₂, 1.8 mM CaCl₂ and 0.2 mM EGTA to approximately 400 µg protein per millilitre. The membrane homogenates were incubated with increasing concentrations of GABA in the presence of 10 µM GDP. [³⁵S]GTPγS (1,250 Ci mmol^{−1}) was then added to a final concentration of 0.5 nM. After incubation at room temperature for 45 min, unbound [³⁵S]GTPγS was removed by centrifugation. The amount of bound [³⁵S]GTPγS was measured using a Beckman LS6500 liquid scintillation counter. Nonspecific binding was measured in the presence of 20 µM unlabelled GTPγS. Basal activity was determined in the absence of GABA. The basal activity of the wild-type receptor was used to calculate the percentage stimulation of the double cysteine mutant receptor GBR1b-Thr198Cys GBR2-Gln206Cys under non-reducing conditions. The reduced [³⁵S]GTPγS binding activity of the double cysteine mutant (~60% of the wild-type value) could be attributed to the effect of the mutations themselves, because introduction of a single cysteine mutation into either subunit also caused a decrease in agonist response.

To measure [³⁵S]GTPγS binding under reducing conditions, the membrane homogenates were pre-incubated with 1 mM DTT before the addition of various concentrations of GABA and 10 µM GDP. The presence of DTT reduced the basal activity of all different combinations of wild-type and cysteine mutant receptors. The percentage stimulation of each receptor mutant was calculated on the basis of the wild-type response obtained under the same condition. Data analysis was performed using the nonlinear regression algorithms in PRISM (GraphPad Software). Data points represent average ± s.e.m. of triplicate measurements.

Radioligand binding assay. HEK293 T/17 cells were transiently transfected with full-length GBR1b and GBR2 plasmids. Cell membranes were suspended in an assay buffer containing 20 mM Tris, pH 7.4, 118 mM NaCl, 5.6 mM glucose, 1.2 mM KH₂PO₄, 1.2 mM MgSO₄, 4.7 mM KCl and 1.8 mM CaCl₂ to approximately 400 µg protein per millilitre. [³H]CGP54626_{ANT} (25 Ci mmol^{−1}) was added to the reaction mixture to final concentrations ranging from 0.5 to 20 nM. After incubation at room temperature for 30 min, unbound [³H]CGP54626_{ANT} was removed by centrifugation. The amount of bound [³H]CGP54626_{ANT} was measured by liquid scintillation counting. Nonspecific binding was measured in the presence of 10 mM unlabelled GABA. Data analysis was performed using the nonlinear regression algorithms in PRISM. Data points represent average ± s.e.m. of triplicate measurements.

Disulphide design and western-blot analysis. The structure of (R)-baclofen_{AGO}-GBR1b_{VFT}:GBR2_{VFT} was used for the rational design of disulphide bonds at the LB2-LB2 heterodimer interface. The residue pair, GBR1b-Thr198 and GBR2-Gln206, was identified by the software DISULPHIDE BY DESIGN⁶² to have the proximity and geometry required for disulphide formation when mutated to cysteines.

The Thr198Cys and Gln206Cys mutations were engineered into full-length GBR1b and GBR2 in pcDNA3.1(+), respectively.

HEK293 T/17 cells were transiently transfected with equal amounts of the full-length GBR1b and GBR2 plasmids. Cells were harvested in a buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl and 1% dodecyl-maltoside. After the insoluble materials were removed by centrifugation, the supernatant was analysed by 4–15% SDS polyacrylamide gel electrophoresis in the absence and presence of 100 mM DTT. In addition, formation of a disulphide-linked heterodimer between the cysteine mutant pair GBR1b-Thr198Cys and GBR2-Gln206Cys was analysed under two different conditions: in the absence of any ligand, and in the presence of 10 mM GABA. Heterodimer formation of all other samples was analysed in the presence of 10 mM GABA. The samples were transferred to polyvinylidene fluoride membranes. After blocking with 5% milk, the membranes were incubated with a primary antibody. Mouse anti-Flag M1 antibody (Sigma) was used to detect the GBR1b protein. Mouse anti-HA antibody HA.11 clone 16B12 (Covance) was used to probe GBR2. Both were followed by an alkaline phosphatase (AP)-conjugated anti-mouse secondary antibody. Proteins were visualized by the colorimetric method.

50. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnol.* **22**, 1583–1587 (2004).
51. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
52. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
53. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
54. Kerr, D. I., Ong, J., Doolette, D. J., Schafer, K. & Prager, R. H. The (S)-enantiomer of 2-hydroxysaclofen is the active GABAB receptor antagonist in central and peripheral preparations. *Eur. J. Pharmacol.* **287**, 185–189 (1995).
55. Frydenvang, K. et al. GABAB antagonists: resolution, absolute stereochemistry, and pharmacology of (R)- and (S)-phaclofen. *Chirality* **6**, 583–589 (1994).
56. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
57. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
58. Roversi, P., Blanc, E., Vonnrhein, C., Evans, G. & Bricogne, G. Modelling prior distributions of atoms for macromolecular refinement and completion. *Acta Crystallogr. D* **56**, 1316–1323 (2000).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Novotny, M., Madsen, D. & Kleywegt, G. J. Evaluation of protein fold comparison servers. *Proteins* **54**, 260–270 (2004).
61. Morin, A. et al. Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).
62. Dombkowski, A. A. Disulfide by Design: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics* **19**, 1852–1853 (2003).

Baryons in the relativistic jets of the stellar-mass black-hole candidate 4U 1630–47

María Díaz Trigo¹, James C. A. Miller-Jones², Simone Migliari³, Jess W. Broderick⁴ & Tasso Tzioumis⁵

Accreting black holes are known to power relativistic jets, both in stellar-mass binary systems and at the centres of galaxies. The power carried away by the jets, and, hence, the feedback they provide to their surroundings, depends strongly on their composition. Jets containing a baryonic component should carry significantly more energy than electron-positron jets. Energetic considerations^{1,2} and circular-polarization measurements³ have provided conflicting circumstantial evidence for the presence or absence of baryons in jets, and the only system in which they have been unequivocally detected is the peculiar X-ray binary SS 433 (refs 4, 5). Here we report the detection of Doppler-shifted X-ray emission lines from a more typical black-hole candidate X-ray binary, 4U 1630–47, coincident with the reappearance of radio emission from the jets of the source. We argue that these lines arise from baryonic matter in a jet travelling at approximately two-thirds the speed of light, thereby establishing the presence of baryons in the jet. Such baryonic jets are more likely to be powered by the accretion disk⁶ than by the spin of the black hole⁷, and if the baryons can be accelerated to relativistic speeds, the jets should be strong sources of γ -rays and neutrino emission.

As part of a campaign to study the connection between relativistic jets and disk winds in X-ray binaries, we used the European Space Agency's XMM-Newton spacecraft and the Australia Telescope Compact Array (ATCA) to make two quasi-simultaneous observations (spaced by less than two days) of the stellar-mass black-hole candidate X-ray binary 4U 1630–47 during its 2012 outburst (Extended Data Table 1). The first X-ray observation, on 2012 September 11–12, showed an X-ray spectrum fully consistent with emission from a standard accretion disk. The quasi-simultaneous radio observation did not detect the source down to 3σ limits on the specific intensity of 68 μ Jy per beam solid angle (angular-resolution element), in images obtained at both 5.5 and 9.0 GHz. In the second observation, on September 28, an additional component was required to model the X-ray spectrum: a non-thermal power law, thermal bremsstrahlung or a Comptonization component (Extended Data Table 2). Furthermore, three narrow X-ray emission lines were significantly detected at energies of 4.04, 7.28 and 8.14 keV (Fig. 1). The quasi-simultaneous ATCA observations detected a radio source at levels of 110 ± 17 and 66 ± 28 μ Jy per beam at 5.5 and 9.0 GHz, respectively.

The strongest emission line, at 7.28 keV, is narrow, with a width of 0.17 ± 0.05 keV. There are no known lines with this rest energy and the most plausible explanation is that this is blueshifted emission from highly-ionized Fe. However, the narrow linewidth, strong blueshift and the lack of an extended red wing (excess emission at long wavelength) in the line profile (which would have indicated relativistic effects) argue against a disk reflection line similar to those previously observed in both X-ray binaries and active galactic nuclei⁸. Although narrow emission lines are frequently observed from the corona in very high-inclination X-ray binaries (the so-called accretion disk corona sources), they are not significantly blueshifted (Supplementary Information). Given the reactivation of the jets implied by the onset of radio emission, we are

left with the intriguing possibility that the 7.28-keV emission line arises from a relativistic jet moving towards the observer.

Taken in isolation, this would constrain the jet velocity to be $>0.3c$, where c is the speed of light, and the inclination angle to be $<73^\circ$ relative to the line of sight, assuming Fe XXVI K α emission ($>0.4c$ and $<66^\circ$, respectively, for Fe XXV K α emission). However, if the 4.04-keV line is associated with the corresponding redshifted Fe XXVI K α emission from the receding jet, we can uniquely constrain the jet velocity to be $0.66c$ and the inclination angle of the jet axis relative to the line of sight to be 65° ($0.63c$ and 63° , respectively, for Fe XXV K α emission). Given that the disk normal is constrained, by the observed X-ray dips and the absence of eclipses⁹, to be in the range 60 – 75° relative to the line of sight, and assuming that the outer and inner disks have the same inclination (that is, there is no warp), this suggests that the jets are aligned with the disk normal, and are hence perpendicular to the disk plane. Furthermore, by assuming identical blueshifts for the 7.28- and 8.14-keV lines, we find a unique identification of such lines as Fe XXVI and Ni XXVII and can self-consistently fit the observed emission lines and the hard continuum component as an emission spectrum from hot, diffuse, Doppler-shifted gas at a temperature of 21 ± 4 keV (energy and temperature being linearly related by Boltzmann's constant), which contributes 19% of the total unabsorbed flux in the 2–10 keV band (Fig. 2).

Fitting a standard multicolour disk and power law with a photon index of $\Gamma = 2$ to model the continuum flux in the second observation gave a bolometric flux in excess of 5.2×10^{-8} erg cm⁻² s⁻¹. However, the exact value depends significantly on the relatively poorly constrained photon index. For a photon index of $\Gamma = 2$ the power-law component comprised 50% of the total bolometric luminosity, whereas for $\Gamma = 2.5$ the power-law contribution was as high as 90%. In either case, this implies that the source was in the 'anomalous' accretion state¹⁰, characterized by a luminosity in excess of 2.5×10^{38} erg s⁻¹ with a dominant contribution to the spectrum from a steep power-law component. The accretion flow is believed to consist of a standard disk, but with a hot corona responsible for Compton upscattering a significant fraction of the disk photons (that is, the photons gain energy after inelastic scattering with energetic electrons in the corona). The only previous detection of jets from 4U 1630–47 also occurred while it was in an 'anomalous' state, when highly polarized, optically thin radio emission was detected¹¹, although no high-resolution X-ray spectra from that outburst are available to search for Doppler-shifted line emission. Contrasting the accretion flow geometry with that of the standard soft state in which jets are quenched by factors of at least several hundred¹² then suggests that it is the presence of the corona, rather than the absence of a standard disk, that is responsible for the launching of the jets.

Narrow, Doppler-shifted X-ray emission lines have previously been reported during an 'apparently standard', slim disk state of 4U 1630–47 observed by NASA's Rossi X-Ray Timing Explorer¹³. However, the poorer spectral resolution of that spacecraft relative to XMM-Newton prevented the definitive association of these lines with either the accretion disk or the jet. Although we find that interpreting these lines as red- and

¹European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München, Germany. ²International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. ³Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain. ⁴School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK. ⁵Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 1710, Australia.

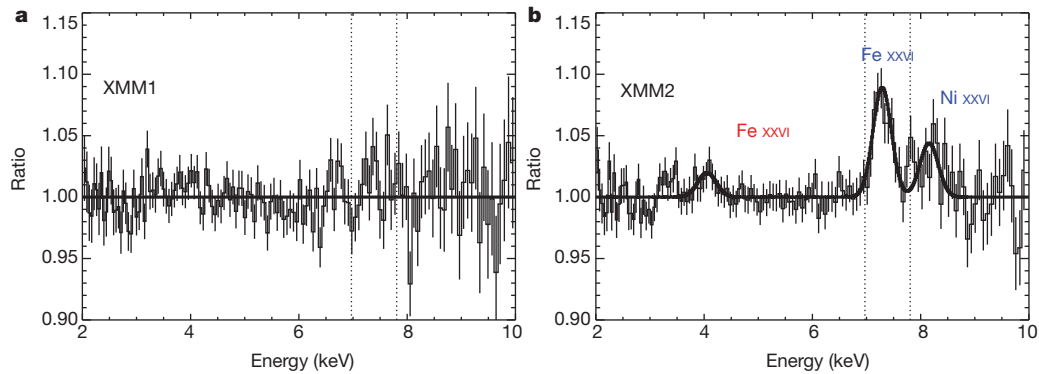


Figure 1 | Residuals from the continuum modelling of the X-ray spectra. Ratio of data (90% error bars) to results of continuum model for the first (a, XMM1) and second (b, XMM2) XMM-Newton observations. The dotted vertical lines indicate the rest energy of the transitions of Fe xxvi (6.97 keV) and Ni xxvi (7.74 keV). The flux ratio between the blue- and redshifted

blueshifted Fe xxvi K α emission from a bipolar jet would give a consistent inclination angle of 58–67° and a slightly lower jet speed of 0.3c–0.4c, the spectral resolution is insufficient to draw more concrete conclusions. Therefore, until now, the only X-ray binary for which there was unambiguous evidence for baryons in the jets was SS 433, in which Doppler-shifted emission lines are seen in both the optical and X-ray bands^{4,5}. However, its persistent, supercritical accretion rate makes it unclear how that system relates to other, more canonical X-ray binaries with significantly more relativistic jets. Although baryons could be supplied to the jets of SS 433 by its strong stellar wind, this mechanism cannot be invoked in other, lower-mass X-ray binaries.

4U 1630–47 is a recurrent transient system. The X-ray spectral and timing features observed during its many well-studied outbursts are typical of other low-mass X-ray binaries, and, together with the absence of type I X-ray bursts, make it one of the best candidates to contain a black hole¹⁴. However, the high column density in the direction of the system has precluded spectroscopic classification of the detected infrared counterpart¹⁵, and a dynamical mass estimate is still lacking. The donor star is most likely to be a relatively early-type (late B or F class) star,

components of Fe xxvi ranges from 1.9 ± 1.1 to 2.1 ± 1.3 (Extended Data Table 3), consistent with 3.2, the ratio predicted for Doppler boosting in a continuous jet. Assuming that the lines are Doppler-broadened by divergence in a conical outflow^{29,30}, we use their widths to determine an upper limit to the opening angle of the jet of 3.7–4.5°.

similar to 4U 1543–47, GRO J1655–40 or SAX J1819.3–2525, with accretion occurring by means of Roche lobe overflow¹⁵. One possible explanation for the non-detection of lines from other systems is that relativistic expansion of standard X-ray binary jets would sufficiently Doppler-broaden any emission lines to the point that their detection would be rendered extremely difficult, even if one knew the true jet velocity and inclination angle, and, hence, the expected redshifts¹⁶. The observed emission lines in 4U 1630–47 could then arise from the particular characteristics of the jets during the poorly studied ‘anomalous’, high-luminosity state in which they were observed.

Most previous attempts to constrain jet composition in both X-ray binaries and active galactic nuclei have relied on energetics considerations, because baryon-loaded jets can carry significant kinetic power away from the compact central objects without radiating. In some active galactic nuclei, the detection of circular polarization has been used to determine the low-energy electron population, and hence to claim, on energetic grounds, that a significant baryonic component can be ruled out³. However, the few reported circular-polarization detections in X-ray binaries¹⁷ were unable to place strong constraints on the jet composition.

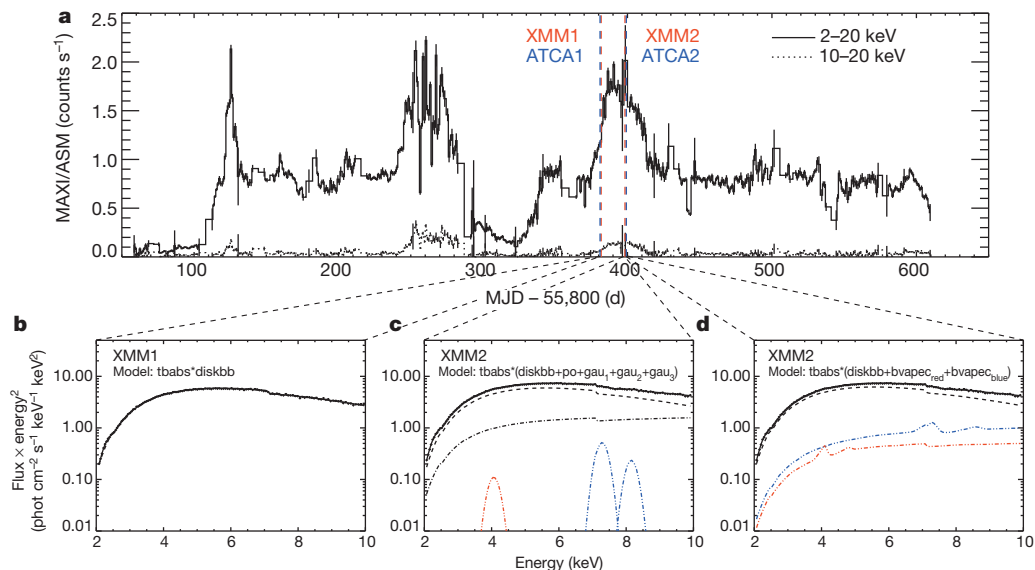


Figure 2 | X-ray observations of 4U 1630–47. a, MAXI/All-Sky Monitor light curve of 4U 1630–47 (MJD, modified Julian date). The times of the XMM-Newton and ATCA observations reported here are indicated with red and blue dashed lines, respectively (Extended Data Table 1). b–d, 2–10-keV unfolded X-ray spectra (flux multiplied by squared energy, shown as a function of energy; error bars (90%) are too small to see). The first observation can be fully described by a standard disk (b). As the source brightens to more

than $2.5 \times 10^{38} \text{ erg s}^{-1}$ in the 2–10-keV band, the spectrum requires an additional hard, power-law, component (black dot-dash line) and three narrow emission lines (red and blue dot-dash lines) (c). Alternatively, two thermal plasma components (bvapec in XSPEC) with a temperature of 21 keV (red and blue dot-dash lines) can account self-consistently for both the hard component and the narrow emission lines (d). Model information refers to XSPEC components.

Arguments based on pressure balance and minimum-energy calculations in the lobes of radio galaxies have suggested that cold protons may carry the bulk of the kinetic energy¹, but there are caveats to these interpretations and uncertainty in the jet composition remains¹⁸. In the best-constrained X-ray binary system, calorimetry of the jet-blown bubble around Cygnus X-1 suggested that the jets should carry a significant cold proton component², although alternative explanations were possible¹⁹. The detection of baryons in the jets of 4U 1630-47 has finally confirmed this picture, at least for certain accretion states.

If they can be accelerated to mildly relativistic speeds, the presence of baryons in an X-ray binary jet suggests that γ -rays could be produced by collisions with high-energy photons or with protons in the stellar wind of the companion star^{20,21}. Such hadronic mechanisms are in principle capable of explaining the observed γ -ray flux from the microquasar Cygnus X-3^{22,23}, although leptonic models seem equally viable²⁴. Even for low-mass X-ray binaries with no strong stellar wind, hadronic models suggest that the presence of relativistic baryons would give rise to γ -ray emission that could be detected by Fermi-LAT, MAGIC II and the CTA²¹. In that case, the hadronic mechanism should also generate an intense flux of neutrinos²⁵. Thus, baryonic jets also have important implications for current and future neutrino telescopes²⁶, and our results suggest that high-luminosity outbursts could provide the best opportunities for neutrino detection.

Finally, the jet composition should be affected by the physical mechanism responsible for launching the jets. Jets powered by an accretion disk⁶ are expected to contain baryons, whereas in the absence of entrainment (as is probably the case in the absence of a strong stellar wind), jets powered by black-hole spin⁷ are more likely to produce purely leptonic jets with significantly higher Lorentz factors. Although there are claims that transient jets from X-ray binaries are powered by black-hole spin²⁷, this work remains controversial²⁸, and so additional evidence detailing the jet composition can provide independent constraints on the jet-launching mechanism.

METHODS SUMMARY

The X-ray observations were made with XMM-Newton, using the EPIC pn charge-coupled-device camera in burst mode. Data were reduced using the XMM-Newton Science Analysis Software. We corrected for a rate-dependent charge transfer inefficiency and re-binned the data, but did not perform background subtraction, owing to both the source brightness and the lack of source-free background regions in burst mode. The resulting X-ray spectra were fitted using the spectral analysis package XSPEC. Whereas the spectrum from 2012 September 11 could be well fitted with a standard disk model, that from September 28 required an additional hard component (Extended Data Table 2). After the spectral continuum had been modelled, a strong, narrow emission feature remained in the residuals of the observation of September 28, at an energy of ~ 7.3 keV. Additional, weaker features were detected at ~ 4.1 and ~ 8.2 keV, whose high statistical significance was confirmed both by an F -test and by Monte Carlo simulations. We also verified the lines to be robust against systematic shifts of up to 2% in the energy scale (although the model became degenerate if the line energy was shifted to coincide with the neutral Fe edge). The best-fitting model parameters are given in Extended Data Table 3.

Quasi-simultaneous radio observations were made with ATCA in its compact H214 configuration, using the Compact Array Broadband Backend to observe simultaneously at 5.5 and 9.0 GHz. Data were initially imported into MIRIAD, and further processed within the Common Astronomy Software Application, using standard procedures. PKS B1934-638 was used as the amplitude and band-pass calibrator, and PMN J1603-4904 was used as the secondary calibrator. No radio emission was detected on 2012 September 10, with an upper limit of three times the root mean squared noise level. The steep-spectrum emission seen on September 29 was fitted with an elliptical Gaussian in the image plane.

Full Methods and any associated references are available in the online version of the paper.

Received 1 May; accepted 16 September 2013.

Published online 13 November 2013.

1. Dunn, R. J. H., Fabian, A. C. & Taylor, G. B. Radio bubbles in clusters of galaxies. *Mon. Not. R. Astron. Soc.* **364**, 1343–1353 (2005).

2. Gallo, E. *et al.* A dark jet dominates the power output of the stellar black hole Cygnus X-1. *Nature* **436**, 819–821 (2005).
3. Wardle, J. F. C., Homan, D. C., Ojha, R. & Roberts, D. H. Electron-positron jets associated with the quasar 3C279. *Nature* **395**, 457–461 (1998).
4. Margon, B., Grandi, S. A., Stone, R. P. S. & Ford, H. C. Enormous periodic Doppler shifts in SS 433. *Astrophys. J.* **233**, L63–L68 (1979).
5. Kotani, A. *et al.* Discovery of the double Doppler-shifted emission-line systems in the X-ray spectrum of SS 433. *Publ. Astron. Soc. Jpn* **46**, L147–L150 (1994).
6. Blandford, R. D. & Payne, D. G. Hydromagnetic flows from accretion discs and the production of radio jets. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
7. Blandford, R. D. & Znajek, R. L. Electromagnetic extraction of energy from Kerr black holes. *Mon. Not. R. Astron. Soc.* **179**, 433–456 (1979).
8. Miller, J. M. Relativistic X-ray lines from the inner accretion disks around black holes. *Annu. Rev. Astron. Astrophys.* **45**, 441–479 (2007).
9. Kuulkers, E. *et al.* Absorption dips in the light curves of GRO J1655–40 and 4U 1630–47 during outburst. *Astrophys. J.* **494**, 753–758 (1998).
10. Abe, Y., Fukazawa, Y., Kubota, A., Kasama, D. & Makishima, K. Three spectral states of the disk X-ray emission of the black-hole candidate 4U 1630–47. *Publ. Astron. Soc. Jpn* **57**, 629–641 (2005).
11. Hjellming, R. M. *et al.* Radio and X-ray observations of the 1998 outburst of the recurrent X-ray transient 4U 1630–47. *Astrophys. J.* **514**, 383–387 (1999).
12. Russell, D. M. *et al.* Testing the jet quenching paradigm with an ultradeep observation of a steadily soft state black hole. *Astrophys. J.* **739**, L19 (2011).
13. Cui, W., Chen, W. & Zhang, S. N. Evidence for Doppler-shifted iron emission lines in black hole candidate 4U 1630–47. *Astrophys. J.* **529**, 952–960 (2000).
14. McClintock, J. E. & Remillard, R. A. in *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) 157–213 (Cambridge Univ. Press, 2006).
15. Augusteijn, T., Kuulkers, E. & van Kerkwijk, M. H. The IR counterpart of the black-hole candidate 4U 1630–47. *Astron. Astrophys.* **375**, 447–454 (2001).
16. Mirabel, I. F., Bandyopadhyay, R., Charles, P. A., Shahbaz, T. & Rodriguez, L. F. The superluminal source GRS 1915+105: a high mass X-ray binary? *Astrophys. J.* **477**, L45–L48 (1997).
17. Fender, R. P. *et al.* Variable circular polarization associated with relativistic ejections from GRS 1915 + 105. *Mon. Not. R. Astron. Soc.* **336**, 39–46 (2002).
18. Worrall, D. The X-ray jets of active galaxies. *Astron. Astrophys. Rev.* **17**, 1–46 (2009).
19. Heinz, S. Composition, collimation, contamination: the jet of Cygnus X-1. *Astrophys. J.* **636**, 316–322 (2006).
20. Romero, G. E., Torres, D. F., Kaufman Bernadó, M. M. & Mirabel, I. F. Hadronic gamma-ray emission from windy microquasars. *Astron. Astrophys.* **410**, L1–L4 (2003).
21. Vila, G. S., Romero, G. E. & Casco, N. A. An inhomogeneous lepto-hadronic model for the radiation of relativistic jets: application to XTE J1118+480. *Astron. Astrophys.* **538**, A97 (2012).
22. Tavani, M. *et al.* Extreme particle acceleration in the microquasar Cygnus X-3. *Nature* **462**, 620–623 (2009).
23. The Fermi LAT Collaboration. Modulated high-energy gamma-ray emission from the microquasar Cygnus X-3. *Science* **326**, 1512–1516 (2009).
24. Piano, G. *et al.* The AGILE monitoring of Cygnus X-3: transient gamma-ray emission and spectral constraints. *Astron. Astrophys.* **545**, A110 (2012).
25. Levinson, A. & Waxman, E. Probing microquasars with TeV neutrinos. *Phys. Rev. Lett.* **87**, 171101 (2001).
26. Aiello, S. *et al.* Sensitivity of an underwater Čerenkov km³ telescope to TeV neutrinos from Galactic microquasars. *Astropart. Phys.* **28**, 1–9 (2007).
27. Narayan, R. & McClintock, J. E. Observational evidence for a correlation between jet power and black hole spin. *Mon. Not. R. Astron. Soc.* **419**, L69–L73 (2012).
28. Russell, D. M., Gallo, E. & Fender, R. P. Observational constraints on the powering mechanism of transient relativistic jets. *Mon. Not. R. Astron. Soc.* **431**, 405–414 (2013).
29. Begelman, M. C., Hatchett, S. P., McKee, C. F., Sarazin, C. L. & Arons, J. Beam models for SS 433. *Astrophys. J.* **238**, 722–730 (1980).
30. Marshall, H. L., Canizares, C. R. & Schulz, N. S. The high resolution X-ray spectrum of SS 433 using the Chandra HETGS. *Astrophys. J.* **564**, 941–952 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was based on observations obtained with XMM-Newton, a European Space Agency (ESA) science mission with instruments and contributions directly funded by ESA member states and the USA (NASA). We thank the XMM-Newton team for the fast scheduling of these observations and the EPIC calibration team for advice. ATCA is part of the Australia Telescope National Facility which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. This work was supported by the Australian Research Council's 'Discovery Projects' funding scheme (J.C.A.M.-J.; project number DP120102393), the Spanish Ministerio de Economía y Competitividad and European Social Funds through a Ramón y Cajal Fellowship (S.M.) and the Spanish Ministerio de Ciencia e Innovación (S.M.; grant AYA2010-21782-C03-01). M.D.T. thanks A. Maury, A. Zdziarski and C. Done for discussions, in which regard J.C.A.M.-J. thanks M. Middleton and S.M. thanks G. Romero, V. Bosch-Ramon, G. Miniutti and S. Motta.

Author Contributions M.D.T., S.M. and J.C.A.M.-J. had the idea for and designed the observing programme. M.D.T. and S.M. analysed the XMM-Newton observations. J.W.B. and T.T. made the radio observations, which were reduced by J.C.A.M.-J. J.C.A.M.-J. and M.D.T. wrote the manuscript, with help from S.M.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D.T. (mdiaztri@eso.org).

METHODS

XMM-Newton. In this section, we provide additional information about the analysis of the XMM-Newton observations of 4U 1630–47. A summary of the observations is shown in Extended Data Table 1. Owing to the high count rate from the source, $\geq 1,000$ counts s^{-1} , we used the EPIC pn charge-coupled-device (CCD) camera in burst mode, in which only one CCD chip is operated and the data are collapsed into a one-dimensional row ($4.4'$) and read out at high speed, the second dimension being replaced by timing information, and applied the Science Analysis Software task *epfast* to the event files to correct for a charge transfer inefficiency effect that has been observed in this mode when high count rates are present³¹. We did not use the EPIC MOS CCD cameras during the observations, to avoid telemetry overflows. Data products were reduced using the Science Analysis Software version 12.0.1. We re-binned the EPIC pn spectra to oversample the full-width at half-maximum of the energy resolution by a factor of three and to have a minimum of 25 counts per bin, to allow the use of the χ^2 statistic. To account for systematic effects, we added a 0.8% uncertainty to each spectral bin after re-binning. We used the pn spectra between 2 and 10 keV.

In the EPIC pn burst mode, there are no source-free background regions, because the point spread function of the telescope extends farther than the central CCD boundaries^{32,33}. Because 4U 1630–47 is very bright, its spectrum will not be significantly modified by the ‘real’ background, which contributes less than 1% to the total count rate across most of the energy band. Conversely, subtracting the background extracted from the outer columns of the central CCD will modify the source spectrum, because the point spread function is energy dependent and the source photons scattered to the outer columns do not show the same energy dependence as the photons focused on the inner columns. Therefore, we checked the effect of subtracting the ‘background’ extracted from the outer regions of the central CCD on the parameters of the lines. Having established that this resulted in no significant changes to the lines, we chose not to subtract such a background when making the final fits, to provide the best possible measurement of the true source spectrum.

We fitted the XMM-Newton data using the spectral analysis package XSPEC³⁴, testing standard X-ray binary models for single-disk emission (*diskbb* in XSPEC), power-law emission (*po* in XSPEC), bremsstrahlung (*bremss* in XSPEC), thermal Comptonization (*comptt* in XSPEC), disk emission plus power-law (*diskbb+po*), disk emission plus bremsstrahlung (*diskbb+bremss*) and disk emission plus thermal Comptonization (*diskbb+comptt*). In each case, we included a neutral absorber (*tbabs* in XSPEC) to account for interstellar absorption along the line of sight. We also included a narrow emission feature at 2.28 keV to account for residual calibration uncertainties at the gold edge of the pn camera³¹.

We found that the first observation could be well fitted by a standard disk (reduced- $\chi^2 = 0.97$ for 128 degrees of freedom (d.o.f.)). In contrast, for the second observation, either a one-component thermal Comptonization model or a two-component model was required to obtain an acceptable fit (reduced- $\chi^2 \leq 2$). A summary of the models tested and the corresponding quality of the fits is shown in Extended Data Table 2. The main difference between the first and second observations is the appearance of a hard component in the latter observation in addition to the thermal disk component. Further support for the existence of this component is given by the increase in the hard-X-ray flux (≥ 10 keV) between the first and second observations, as detected by MAXI/All-Sky Monitor and Swift/BAT. In particular, the Swift/BAT 15–50-keV count rate increased from 0.019 ± 0.001 counts $cm^{-2} s^{-1}$ to 0.036 ± 0.002 counts $cm^{-2} s^{-1}$. However, we cannot discriminate between the three components (power law, bremsstrahlung or thermal Comptonization) used to model the additional emission in the second observation, owing to the low effective area of XMM-Newton at energies > 10 keV, at which such a component should become dominant. We found that a one-component, thermal Comptonization model with low seed photon and plasma temperatures (~ 0.7 keV and 1.9 keV, respectively) and a high optical depth ($\tau \approx 8$) could fit the second observation just as well as the two-component models. However, the 15–50-keV flux predicted by the one-component model, $\sim 3 \times 10^{-10}$ erg $cm^{-2} s^{-1}$, is lower than the flux detected by Swift/BAT, $\sim 2 \times 10^{-9}$ erg $cm^{-2} s^{-1}$, by almost one order of magnitude. In contrast, the two-component models predict a 15–50-keV flux of $(\sim 2\text{--}3) \times 10^{-9}$ erg $cm^{-2} s^{-1}$, in excellent agreement with Swift/BAT. Therefore, we discard the one-component thermal Comptonization model for the rest of this work, but include it for completeness in Extended Data Tables 2 and 3, together with all the other tested models.

Having modelled the continuum emission, a strong, narrow emission feature remained in the residuals of the second observation at ~ 7.3 keV, with additional weaker features at energies of ~ 3.5 , 4.1, 7.8 and 8.2 keV. Inclusion of a narrow Gaussian at ~ 7.28 keV, to account for the most significant feature, improved the fit from $\chi^2_{\nu} = 1.70$ (127 d.o.f.) to 1.07 (124 d.o.f.). The probability of such an improvement occurring by chance, as indicated by an *F*-test, is 5×10^{-13} . We also tested the significance of the weaker features, including them one by one in addition to

the feature at 7.28 keV. Because the weak features are narrow, we coupled their widths to that of the 7.28-keV feature to prevent an arbitrary increase in the linewidths from absorbing deficiencies in the continuum model. We found that independently including the ~ 4.1 -keV and ~ 8.2 -keV features improved the quality of the fit to $\chi^2 = 0.97$ (122 d.o.f.) and 0.95 (122 d.o.f.), respectively (an *F*-test indicated that the probabilities of such improvements occurring by chance were 0.0007 and 0.0002, respectively). For the feature at 3.5 keV, the chance probability of the fit improvement as given by the *F*-test was 0.023. The feature at 7.8 keV is weaker than that at 8.2 keV, and unless the energy was fixed, the fit shifted the energy to 8.2 keV. If the feature at 7.8 keV was included in addition to the 8.2-keV feature, the chance probability of the fit improvement as given by the *F*-test was 0.041.

Because there are caveats to the use of the *F*-test to study the presence of emission lines over a continuum³⁵, we also performed Monte Carlo simulations to confirm the significance of the 4.1-keV and 8.2-keV features³⁶. As an example, we took the best-fitting parameters of the model consisting of absorbed disk emission plus thermal Comptonization (*diskbb+comptt*; Extended Data Table 3) and the 7.28-keV Gaussian emission line as our null hypothesis. We then tested the chance probability of any extra emission line against the null hypothesis.

For this, we fitted the data with the null-hypothesis model and simulated a spectrum with the same exposure time as the data. We fitted the simulated spectrum with the model used to construct it, providing a refined null-hypothesis model that differed from the previous one only in photon statistics. We then added to the model a Gaussian emission line, varying its energy from 3.0 to 9.0 keV in steps of 0.2 keV. We chose the linewidth to be equal to the one that we found in our best-fit model of the data with the three Gaussian emission lines, $\sigma = 0.17 \pm 0.05$ keV, and we allowed the normalization to vary. For each spectrum, we recorded the best-fitting parameters and the maximum $\Delta\chi^2$ among the lines. We then repeated the above steps 1,000 times and obtained the distribution of maximum $\Delta\chi^2$ to be compared with the result obtained from the data. The addition of an extra Gaussian to the data at 4.1 or 8.2 keV improved the fit by $\Delta\chi^2 = 12.73$ or 13.77, respectively, compared with the null hypothesis.

Only six and, respectively, one of the 1,000 simulated spectra showed a maximum $\Delta\chi^2$ equal to or higher than those found by fitting the data with the additional 4.1- and 8.2-keV lines. Therefore, these additional lines are significant at levels of 99.4% and 99.9%, respectively.

In summary, given the higher significance of the ~ 4.1 - and ~ 8.2 -keV features, we consider it likely that they are real and we discard the other weak features, which have significantly higher chance probabilities. In Extended Data Table 3, we show the parameters of the best-fit models. The lines are robust against significance tests, and we also tested whether they are robust against possible systematic errors in the energy scale. For this, we shifted the energy scale by as much as $\pm 2\%$ in the event files of both observations and re-extracted spectra and response matrices. Coincidentally, the maximum possible systematic error of $\sim 2\%$ in the energy scale would shift the photons of the line at 7.28 keV to the energy of the neutral Fe edge, causing model degeneracy. Therefore, we encourage future exposures with high-resolution gratings to mitigate against possible systematics and definitively confirm the exact energies of the lines.

Finally, we searched for line variability within the second observation by dividing the events file into two, three and four different parts and extracting and fitting the corresponding spectra. We did not detect any significant variation of the lines among the different intervals. However, owing to the poor statistics we obtain large errors and cannot exclude variability with high significance. Longer exposures with high-resolution gratings would be required to address definitively the existence of line variability, as might be expected in the case of jet precession.

Australia Telescope Compact Array. ATCA was used to observe 4U 1630–47 at two epochs, as detailed in Extended Data Table 1. We made simultaneous observations at 5.5 and 9.0 GHz using the Compact Array Broadband Backend system³⁷. We observed with 2,048 MHz of contiguous bandwidth in each of the two frequency bands, split into 1-MHz channels. The array was in the hybrid H214 configuration, with five of the six antennas within 250 m, and the sixth antenna providing longer (4.5-km) baselines.

Data were imported into MIRIAD³⁸, and immediately written out to FITS format for further processing within the Common Astronomy Software Application³⁹. We edited out narrowband radio-frequency interference, before deriving the external gain calibration, using PKS B1934–638 as the band-pass calibrator and to set the amplitude scale. We used PMN J1603–4904 to derive the amplitude and phase gains that were subsequently applied to the target field. To filter out the diffuse emission in the field and enhance the sensitivity to point sources, we only used the long baselines to antenna 6 in making the radio images. We used natural weighting to maximize the image sensitivity. The source was too faint for self-calibration. On 2012 September 10, the source was not detected at either frequency, and we estimated the 3σ upper limit on its flux density as three times the root mean squared noise in the image. On 2012 September 29, when the source was significantly

detected, we measured its flux density by fitting an elliptical Gaussian to the source in the image plane.

31. Guainazzi, M. *et al.* Evaluation of the spectral calibration accuracy in EPIC-pn fast modes. Technical Note XMM-SOC-CAL-TN-0083; <http://xmm.vilspa.esa.es/docs/documents/CAL-TN-0083.pdf> (2012).
32. Done, C. & Díaz Trigo, M. A re-analysis of the iron line in the XMM-Newton data from the low/hard state in GX 339-4. *Mon. Not. R. Astron. Soc.* **407**, 2287–2296 (2010).
33. Ng, C., Díaz Trigo, M., Cadolle Bel, M. & Migliari, S. A systematic analysis of the broad iron K α line in neutron-star LMXBs with XMM-Newton. *Astron. Astrophys.* **522**, A96 (2010).
34. Arnaud, K. A. in *Astronomical Data Analysis Software and Systems V* (eds Jacoby, G. H. & Barnes, J.) 17 (ASP Conf. Ser. 101, Astronomical Society of the Pacific, 1996).
35. Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L. & Siemiginowska, A. Statistics, handle with care: detecting multiple model components with the likelihood ratio test. *Astrophys. J.* **571**, 545–559 (2002).
36. Miniutti, G. & Fabian, A. C. Discovery of a relativistic Fe line in PG 1425+267 with XMM-Newton and study of its short time-scale variability. *Mon. Not. R. Astron. Soc.* **366**, 115–124 (2006).
37. Wilson, W. E. *et al.* The Australia Telescope Compact Array Broad-band Backend: description and first results. *Mon. Not. R. Astron. Soc.* **416**, 832–856 (2011).
38. Sault, R. J., Teuben, P. J. & Wright, M. C. H. in *Astronomical Data Analysis Software and Systems IV* (eds Shaw, R. A., Payne, H. E. & Hayes, J. J. E.) 433–436 (ASP Conf. Ser. 77, Astronomical Society of the Pacific, 1995).
39. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. in *Astronomical Data Analysis Software and Systems XVI* (eds Shaw, R. A., Hill, F. & Bell, D. J.) 127–130 (ASP Conf. Ser. 376, Astronomical Society of the Pacific, 2007).

Extended Data Table 1 | Observation log

Observation	Observation ID	Observation Times (UTC) (day.month.year hr:min)
XMM1	0670673101	11.09.2012 20:14 – 12.09.2012 05:39
ATCA1	C2514	10.09.2012 01:08 – 10.09.2012 09:03
XMM2	0670673201	28.09.2012 06:33 – 28.09.2012 21:50
ATCA2	C2514	29.09.2012 02:36 – 29.09.2012 09:58

Extended Data Table 2 | Quality of the spectral fits (χ^2_ν) to the XMM-Newton observations, using different continuum models

Model	χ^2_ν (d.o.f.)	
	XMM1	XMM2
tbabs*diskbb	0.97 (128)	2.12 (128)
tbabs*po	9.44 (128)	12.9 (128)
tbabs*bremss	2.59 (128)	3.26 (128)
tbabs*comptt	0.87 (126)	1.26 (126)
tbabs*(diskbb+po)	0.95 (127)	1.70 (127)
tbabs*(diskbb+bremss)	0.96 (126)	1.70 (126)
tbabs*(diskbb+comptt)	0.95 (127)	1.67 (127)

Extended Data Table 3 | Parameters for each of the best-fit models to the 2–10-keV EPIC pn spectra from the XMM-Newton observations

Obs.	Best-fit model									χ^2_{ν} (d.o.f.)
	tbabs*diskbb									
	N_{Habs}	kT_{dbb} (keV)	k_{dbb}							
XMM1	8.34 ± 0.08	1.82 ± 0.01	107 ± 3	–	–	–	–	–	–	0.97 (128)
	tbabs*(comptt+gau+gau+gau)									
	N_{Habs}	kT_{bb} (keV)	kT_e (keV)	τ	k_{comptt}	E_{gau} (keV)	σ (keV)	$Flux_{gau}$	EW (eV)	
XMM1	7.5 ± 0.4	$0.69^{+0.08}_{-0.06}$	1.63 ± 0.07	9.1 ± 0.9	3.0 ± 0.3	–	–	–	–	0.87 (126)
XMM2	$7.36^{+0.09}_{-0.26}$	$0.72^{+0.05}_{-0.02}$	$1.86^{+0.06}_{-0.04}$	7.7 ± 0.5	$3.23^{+0.09}_{-0.22}$	4.06 ± 0.07	(t)	1.9 ± 1.0	5 ± 2	0.79 (119)
						7.24 ± 0.03	0.11 ± 0.03	3.8 ± 0.8	23 ± 4	
						8.22 ± 0.10	(t)	0.9 ± 0.7	7 ± 4	
	tbabs*(diskbb+po+gau+gau+gau)									
	N_{Habs}	kT_{dbb} (keV)	k_{dbb}	Γ	k_{po}	E_{gau} (keV)	σ (keV)	$Flux_{gau}$	EW (eV)	
XMM1	8.5 ± 0.2	1.78 ± 0.03	113 ± 6	2 (f)	< 0.6	–	–	–	–	0.95 (127)
XMM2	8.6 ± 0.1	1.77 ± 0.03	123 ± 7	2 (f)	1.7 ± 0.3	$4.04^{+0.10}_{-0.12}$	(t)	3.0 ± 1.4	8 ± 4	0.87 (120)
						7.28 ± 0.04	0.17 ± 0.05	6.2 ± 1.0	37 ± 9	
						$8.14^{+0.14}_{-0.18}$	(t)	2.5 ± 1.0	18 ± 10	
	tbabs*(diskbb+bremsstrahlung+gau+gau+gau)									
	N_{Habs}	kT_{dbb} (keV)	k_{dbb}	kT_{brems} (keV)	k_{brems}	E_{gau} (keV)	σ (keV)	$Flux_{gau}$	EW (eV)	
XMM1	8.6 ± 0.4	1.79 ± 0.06	102^{+16}_{-24}	> 1.2	$0.9^{+1.8}_{-0.8}$	–	–	–	–	0.95 (126)
XMM2	$8.4^{+0.4}_{-0.2}$	1.74 ± 0.05	134^{+15}_{-25}	> 5.3	$1.10^{+1.74}_{-0.05}$	$4.04^{+0.10}_{-0.12}$	(t)	3.2 ± 1.3	8 ± 4	0.87 (119)
						7.27 ± 0.04	0.17 ± 0.05	6.1 ± 0.9	37 ± 8	
						$8.15^{+0.15}_{-0.18}$	(t)	2.3 ± 0.9	18^{+11}_{-9}	
	tbabs*(diskbb+comptt+gau+gau+gau)									
	N_{Habs}	kT_{dbb} (keV)	k_{dbb}	kT_e (keV)	τ	k_{comptt}	E_{gau} (keV)	σ (keV)	$Flux_{gau}$	EW (eV)
XMM1	8.5 ± 0.1	$1.7^{+0.1}_{-0.3}$	132^{+34}_{-26}	68^{+432p}_{-18p}	< 0.58	< 0.75	–	–	–	0.96 (125)
XMM2	8.3 ± 0.1	1.6 ± 0.2	212^{+60}_{-70}	145^{+355p}_{-95p}	< 0.50	$0.005^{+0.3}_{-0.004}$	$4.03^{+0.11}_{-0.13}$	(t)	2.9 ± 1.4	8 ± 4
							7.27 ± 0.04	0.17 ± 0.05	6.0 ± 1.0	37 ± 8
							$8.15^{+0.14}_{-0.18}$	(t)	2.4 ± 1.0	19 ± 11
	tbabs*(diskbb+bvapec_{red}+bvapec_{blue})									
	N_{Habs}	kT_{dbb} (keV)	k_{dbb}	kT_{bvapec} (keV)	$z_{bvapec,red}$	k_{red}	$z_{bvapec,blue}$	k_{blue}	Ni	v (km/s)
XMM2	8.3 ± 0.1	1.77 ± 0.04	130 ± 10	21 ± 4	0.706 ± 0.026	2.0 ± 0.8	$-0.044^{+0.003}_{-0.010}$	1.3 ± 0.3	< 5.4	5260^{+4850}_{-1730}

Uncertainties are quoted at the 90% level. N_{Habs} is the column density of the neutral absorber in units of 10^{22} cm^{-2} . k_{dbb} , k_{po} , k_{brems} and k_{comptt} are the normalizations of the disk black-body, power-law, bremsstrahlung and Comptonization components in XSPEC units. kT_{dbb} , kT_{brems} and kT_{bvapec} are the temperatures of the disk black-body, bremsstrahlung and bvapec components in units of keV. For the comptt component, we imposed a lower limit on the temperature of the plasma of 50 keV and an upper limit on the opacity of 2, and coupled the temperature of the seed photons to the temperature of the disk black body in the two-component model. The photon index of the power law, Γ , has been fixed to 2, because it is poorly constrained. k_{blue} and k_{red} respectively represent the normalizations of the blue and red bvapec components in XSPEC units. z_{blue} and z_{red} respectively are the velocity shifts for the blue and red bvapec components. 'Ni' represents the nickel abundance with respect to solar abundances; v represents the velocity broadening for the bvapec components, for which we imposed the same value for the red and blue components. E_{gau} , σ , EW and $Flux_{gau}$ respectively represent the energy (in units of keV), width (in units of keV), equivalent width (in units of eV) and 2–10-keV unabsorbed flux (in units of $10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$) of the Gaussian features. The width, σ , was tied for all the emission lines during the fit. 'p' indicates that a parameter was pegged at its lower or upper limit.

An exactly solvable model for quantum communications

Graeme Smith¹ & John A. Smolin¹

Information theory establishes the ultimate limits on performance for noisy communication systems¹. Accurate models of physical communication devices must include quantum effects, but these typically make the theory intractable^{2–5}. As a result, communication capacities—the maximum possible rates of data transmission—are not known, even for transmission between two users connected by an electromagnetic waveguide with Gaussian noise⁶. Here we present an exactly solvable model of communication with a fully quantum electromagnetic field. This gives explicit expressions for all point-to-point capacities of noisy quantum channels, with implications for quantum key distribution and fibre-optic communications. We also develop a theory of quantum communication networks by solving some rudimentary models including broadcast and multiple-access channels. We compare the predictions of our model with the orthodox Gaussian model and in all cases find agreement to within a few bits. At high signal-to-noise ratios, our simple model captures the relevant physics while remaining amenable to exact solution.

A fundamental property of communication systems is the maximum rate of data transmission possible using the best communication schemes. This is called the ‘capacity’ of a channel. It is usually calculated as a function of noise levels and subject to a limited power budget. Shannon¹ presented a theory of information both formulating and solving the capacity problem. For a channel with additive white Gaussian noise, his formula can be solved explicitly, giving the classical capacity, C , as a function of bandwidth, B , signal power, W , and noise power per hertz, N :

$$C = B \log \left(1 + \frac{W}{NB} \right) \quad (1)$$

This formula guided the development of practical schemes now in use, culminating in efficient codes approaching the theoretical limit⁷.

Noise is not a purely mathematical abstraction, but must arise from some physical process. Such processes are properly described by quantum mechanics, and calculating the true information-carrying capacity of a channel therefore requires a quantum mechanical treatment^{4,5}. It is natural, then, to consider new types of capacity, such as the capacity of a channel to transmit quantum states coherently (the ‘quantum capacity’^{8,9}) or classical states securely (the ‘private capacity’⁸). However, unlike in classical information theory, for most quantum channels none of these capacities is known (see, for example, refs 10, 11 regarding quantum capacity, ref. 12 regarding classical capacity and ref. 13 regarding private capacity).

Whereas much existing work on quantum channels concentrates on abstract, finite-dimensional channels, here we study a more realistic setting. Our method is well suited to Gaussian noise in electromagnetic modes, but is substantially more general. We will be able to calculate classical, quantum and private capacities for a wide range of realistic channels.

Because quantum information cannot be cloned¹⁴, knowledge gained by the environment about a signal is necessarily detrimental to quantum transmission. This need to consider information transmitted to the environment as well as to the intended receiver puts an analysis of quantum capacity on a par with the study of the classical multi-user

broadcast channel (also very difficult to analyse¹⁵). To make the problem tractable, we follow ref. 16 in substituting a discretized and deterministic model for the actual channel. In the limit of high signal-to-noise ratio, this model captures the important features of the real channel and allows us to calculate capacities to within a small number of bits. The approach can be thought of as a discretization of the continuous system under consideration in a very simple way and then a truncation of signals smaller than the noise. The result is a deterministic model that is easy to analyse and yields exact answers for the capacities in our model. We present strong evidence that these models give good approximations, but our estimates do not directly give upper or lower bounds on the capacities of the bosonic Gaussian channels being modelled.

States of electromagnetic modes are described by their quadratures (P, Q) and are called Gaussian when they are completely characterized by a matrix containing the covariances of these quadratures¹⁷. Such states can be visualized by their Wigner functions, which are quasi-probability distributions depicting the state’s location in phase space. Gaussian states have ellipsoidal Wigner functions with Gaussian profiles. Figure 1a shows the Wigner function for a Gaussian state with variances σ_P^2 and σ_Q^2 . The minimum-uncertainty state has $\sigma_P\sigma_Q = 1/2$ (in units of \hbar , which is Planck’s constant divided by 2π). Such states are always pure. Mixed states have $\sigma_P\sigma_Q > 1/2$ and can be thought of as mixtures of pure states.

We replace the quadratures (P, Q) with a discretized model (Fig. 1b). We call this the discrete quadrature model (DQM). Roughly speaking,

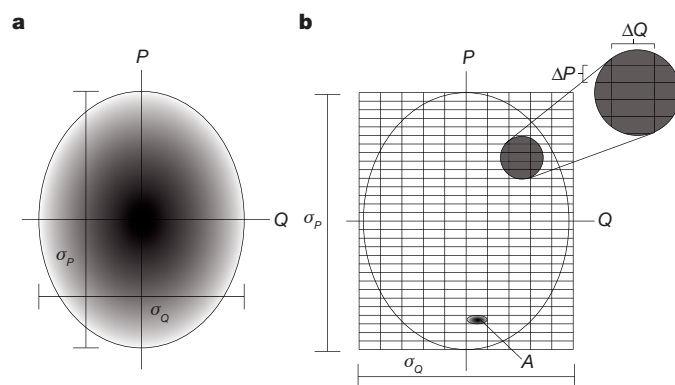


Figure 1 | Phase-space representation of states (Wigner functions).

a, Wigner function of a Gaussian mixed state with variances σ_Q and σ_P . **b**, The discrete quadrature model. The mixed state has been approximately decomposed into non-overlapping rectangles with width ΔQ and height ΔP . These rectangles are the discrete states of the model that replaces the original physical system. We imagine them as approximations to the pure squeezed state shown at point A. These non-overlapping rectangles are perfectly distinguishable in our model. Any inaccuracy in counting the total number of rectangles that results from our replacement of a Gaussian-profiled ellipsoid in phase space with a tiling of a rectangular region of area $\sigma_P\sigma_Q$ will be unimportant because the capacities we study are logarithmic quantities and even a multiplicative factor of two in our counting will change the capacity by only a single bit.

¹IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA.

states in phase space are more distinguishable when their Wigner functions are less overlapping. To reflect this property, we divide phase space into a lattice of non-overlapping rectangles that we take to be perfectly distinguishable. The smallest physical state has area $1/2$. The Wigner function is thus replaced with several rectangles of area $1/2$ that tile the region of phase space where the function is non-negligible. The associated state in our model is a uniform mixture of these distinguishable rectangles.

Our channel model is also discrete: because states in our model are perfectly distinguishable rectangles, every pair of input states will be mapped either to distinguishable outputs or to the same output. Calculating the communication rate is then a matter of counting distinguishable outputs. The capacity is the logarithm of the highest number of distinguishable states achievable with any coding–decoding strategy (logs are to base 2 throughout). An important feature of the capacity being a logarithmic quantity is that we can safely round up whenever a number of states is non-integral because even overcounting by a factor of two would change the calculated capacity by only one bit, and we are not aiming for more accuracy than that. Another reason to round up is to avoid ever computing a capacity of $\log_2 0$, which would result in meaningless infinities.

A full calculation of capacity includes a maximization over modulation schemes, that is, the set of input states employed. To remain physical, the choice of modulation must obey

$$\begin{aligned} (\sigma_Q^2 + \sigma_P^2)/2 &\leq W && \text{(finite power)} \\ \Delta P \Delta Q &\geq 1/2 && \text{(quantum uncertainty)} \\ \Delta P &\leq \sigma_P, \Delta Q \leq \sigma_Q && \text{(common sense)} \end{aligned}$$

Because in our model all output states are either perfectly distinguishable or completely indistinguishable, the decoding schemes are always straightforward and no special care to ensure physicality is required. However, for quantum capacity we will have to consider three possibilities rather than two, namely confusable input states, input states that are distinguishable to the output alone and those that are distinguishable to both the output and the environment.

The evolution of bosonic states under the action of channels is described by the evolution of the system's quadratures. For Gaussian noise, which arises from quadratic interaction with Gaussian environment modes, the allowed evolutions take a particularly simple form, being completely described by a symplectic matrix⁶. For example, $P \rightarrow \sqrt{\lambda}P + \sqrt{1-\lambda}r$, $Q \rightarrow \sqrt{\lambda}Q + \sqrt{1-\lambda}s$ results from interaction with an environment mode described by (r, s) . The ‘additive Gaussian noise channel’ arises when the environment begins in a Gaussian state, and the ‘thermal noise channel’ arises when the initial state is thermal.

We now apply our discretization procedure to classical communication over an additive Gaussian noise channel. Figure 2 illustrates the following analysis. We fix a modulation scheme, to be optimized over later, which amounts to deciding the shape of the rectangles in the discretization shown in Fig. 1b. Given σ_P, σ_Q and rectangle shape $(\Delta P, \Delta Q)$ with $\Delta P \Delta Q = 1/2$, our input space has $\sigma_P \sigma_Q / \Delta P \Delta Q$ distinguishable states (Fig. 2a). We must now determine how many distinguishable outputs these get mapped to. Attenuation by λ maps the entire input space to a $\sqrt{\lambda}\sigma_P \times \sqrt{\lambda}\sigma_Q$ rectangle of $\sqrt{\lambda}\Delta P \times \sqrt{\lambda}\Delta Q$ tiles (Fig. 2b). When noise of typical size $\sqrt{1-\lambda}\sigma_r$ is added to the P quadrature, tiles closer than this are taken to be confusable, and similarly for Q , which results in ‘metatiles’ of dimension (Fig. 2c)

$$\max(\sqrt{\lambda}\Delta P, \sqrt{1-\lambda}\sigma_r) \times \max(\sqrt{\lambda}\Delta Q, \sqrt{1-\lambda}\sigma_s)$$

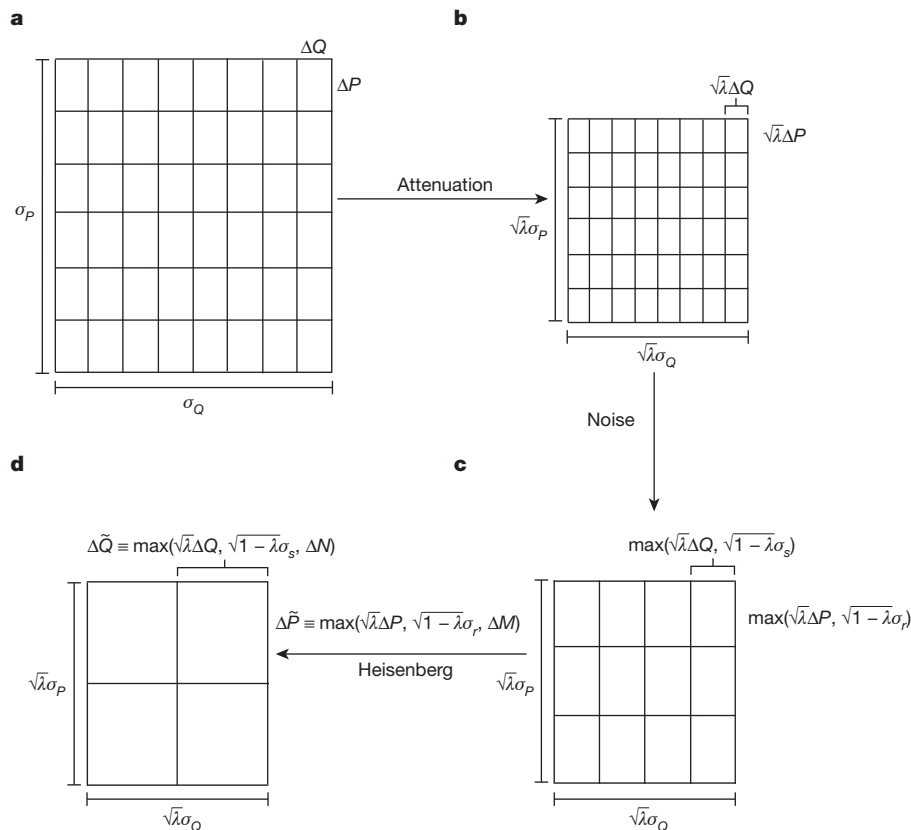


Figure 2 | The DQM of the action of a channel. **a**, The input space of size $\sigma_P \times \sigma_Q$ with power $(\sigma_P^2 + \sigma_Q^2)/2 \leq W$ is divided up into tiles of dimensions $\Delta Q \times \Delta P$. **b**, Attenuation has shrunk the tiles to $\sqrt{\lambda}\Delta Q \times \sqrt{\lambda}\Delta P$. **c**, Noise

renders some tiles confusable, and so the effective tile size is increased. **d**, If any tiles are smaller than the smallest tile size allowed by Heisenberg uncertainty, the tiles are enlarged again.

These new tiles fill up the large $\sqrt{\lambda}\sigma_P \times \sqrt{\lambda}\sigma_Q$ rectangle of Fig. 2b, as pictured in Fig. 2c. If it happens that the tile size is less than $1/2$ then the tiles are smaller than the minimum size allowed by uncertainty. We must then choose a tile shape, $\Delta N \times \Delta M$ with $\Delta N \Delta M = 1/2$, satisfying the uncertainty limit. We thus find a final tile dimension of $\Delta \tilde{P} \times \Delta \tilde{Q}$ with (Fig. 2d)

$$\Delta \tilde{P} \equiv \max(\sqrt{\lambda}\Delta P, \sqrt{1-\lambda}\sigma_r, \Delta N) \quad (2)$$

$$\Delta \tilde{Q} \equiv \max(\sqrt{\lambda}\Delta Q, \sqrt{1-\lambda}\sigma_s, \Delta M) \quad (3)$$

This gives a total number of distinguishable output states of $\lambda\sigma_P\sigma_Q/\Delta\tilde{P}\Delta\tilde{Q}$ and a classical capacity of

$$C = \max \log \left(\frac{\lambda\sigma_P\sigma_Q}{\Delta\tilde{P}\Delta\tilde{Q}} \right)$$

where the maximization is over all of the constrained variables: $(\sigma_P^2 + \sigma_Q^2)/2 \leq W$, $\Delta P \Delta Q \geq 1/2$, $\Delta P \leq \sigma_P$, $\Delta Q \leq \sigma_Q$, $\Delta N \Delta M \geq 1/2$. We note that maximization over ΔP and ΔQ corresponds to a choice of encoding strategy, whereas maximization over ΔM and ΔN corresponds to a choice of measurement by the receiver. We evaluate this formula for some important special cases:

(1) The attenuation channel: an additive channel with transmissivity λ and pure environment in a vacuum state with $\sigma_r = \sigma_t = 1/\sqrt{2}$. To a good approximation, this channel describes propagation in lossy optical fibre. Letting $\Delta P = \Delta Q = 1/\sqrt{2}$ and $\Delta M = \Delta N = 1/\sqrt{2}$ minimizes $\Delta\tilde{P}\Delta\tilde{Q} = 1/2$ while satisfying the constraints. This leads to a capacity of $C = \log(2W\lambda)$. For this channel, we actually know the classical capacity exactly¹⁸: the true capacity, $g(\lambda(W - 1/2))/\ln 2$, where $g(x) = (x+1)\ln(x+1) - x\ln(x)$, differs from our estimate by no more than 1.4 bits (Fig. 3c, d).

(2) A channel that applies Gaussian-distributed translations to both quadratures in phase space (the classical noise channel). This arises as the limiting case of the thermal noise channel with $\lambda \rightarrow 1$, $\sigma_r = \sigma_s \rightarrow \infty$ and $\sigma_r\sqrt{1-\lambda} = \mu$ held constant. We find that

$$C = \min[\log(2W), \log(W/\mu^2)] \quad (4)$$

This is within 2.45 bits of the true capacity (Supplementary Information).

(3) A channel that applies Gaussian-distributed translations to a single quadrature. Such a channel depends on the noise power added, μ^2 , and the capacity might be expected to depend on this parameter. In fact, the DQM of this channel gives a capacity of $C = \log(2W)$, independent of μ^2 (Supplementary Information). Although the predicted capacity is independent of μ^2 , the suggested modulation scheme is not: for $\mu^2 \geq 1$ it consists of squeezed states of width of order $1/2\sqrt{W}$ in the noiseless quadrature, whereas for $\mu^2 \leq 1$ unsqueezed states suffice. All these predictions are borne out by comparison with the standard Gaussian model.

As mentioned above, evaluating the quantum capacity, Q , requires assessing not only signals sent from sender to receiver, but also how information leaks from the channel to the environment. Although this makes our calculations more complex, we nevertheless find simple and reliable estimates for quantum capacities. The quantum capacity is the number of distinguishable states that are communicated to the channel's output and about which the environment knows nothing. This complete lack of knowledge makes it possible to communicate quantum superpositions of these distinguishable states; therefore, in our model we use them to define the basis states of a Hilbert space that will be successfully transmitted. We note that this is exactly the definition of the private capacity⁸, and as a result our model will not be able to distinguish between the quantum and private capacities.

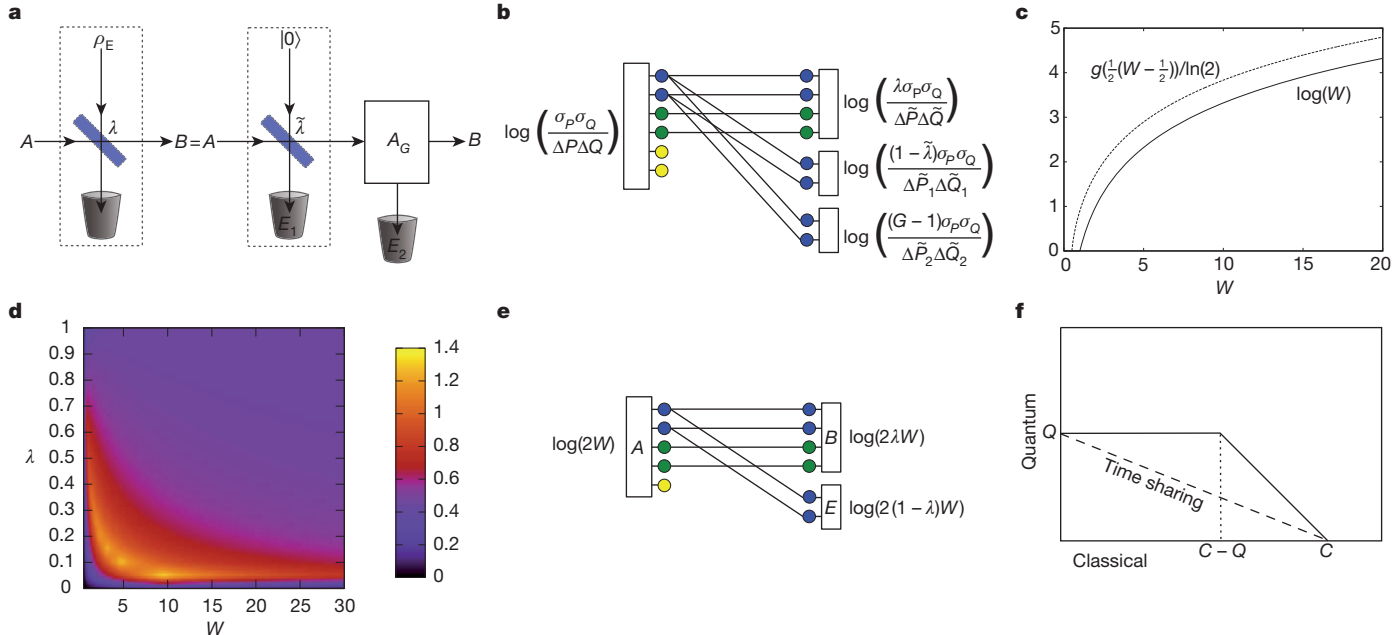


Figure 3 | Attenuation and thermal noise channels. **a**, Additive Gaussian noise channel. An input state in beam A is combined with an environmental input, ρ_E , on a beam splitter with transmissivity λ . One output beam is discarded (remaining available only to the environment) and the other is the channel output, beam B . When the environment's input, ρ_E , is a vacuum state, the channel is an attenuation channel. When ρ_E is a thermal state, the channel is a thermal noise channel. The thermal noise channel with average photon number N_E can be decomposed into a pure attenuation channel with $\tilde{\lambda} = \lambda/G$ and a quantum limited amplifier, A_G , with gain G . The output of the channel is labelled B , and the environments of the attenuation and amplification channel are E_1 and E_2 , respectively. **b**, DQM of a thermal noise channel using this decomposition. **c**, The classical capacity of the attenuation channel for $\lambda = 1/2$.

We have plotted the actual capacity, $g((W - 1/2)/2)/\ln(2)$, and the capacity calculated using the DQM, $\log(W)$. **d**, The difference between the actual and discrete quadrature capacities for the attenuation channel, as a function of both λ and power. The difference is never greater than 1.4 bits. **e**, DQM for the Gaussian attenuation channel. **f**, Rate region for simultaneous communication of classical and quantum information over the Gaussian attenuation channel. The highest-order (blue) bits are transmitted to the environment and therefore can never be used for quantum communication. They can always be used for classical communication, even when the lower-order bits are being used for quantum communication. This offers an advantage over the straight-line trade-off associated with time sharing.

The DQM of the thermal noise channel, including the environment, is shown in Fig. 3b. It consists of a number of input bits, depending on the channel parameters and transmission power, that are transmitted noiselessly to the receiver (green and blue bits). A smaller number of higher-order bits are also copied and sent to the environment (blue). The bits transmitted to the receiver but not the environment (green bits) are transmitted coherently, so that the difference between the respective numbers of bits sent to the receiver and to the environment gives the quantum capacity. The resulting formula (Supplementary Information) is

$$Q = \log(\lambda) - \log\left(1 - \frac{\lambda}{(1-\lambda)N_E + 1}\right) - \log\left(\frac{\Delta\tilde{P}\Delta\tilde{Q}}{\Delta\tilde{P}_1\Delta\tilde{Q}_1}\right) \quad (5)$$

where $\Delta\tilde{P}$ and $\Delta\tilde{Q}$ are given in equations (2) and (3), respectively, N_E is the average number of photons in the environment (Fig. 3a) and

$$\Delta\tilde{P}_1 = \max\left(\sqrt{1-\tilde{\lambda}}\Delta P, \Delta M_1, \sqrt{\tilde{\lambda}/2}\right)$$

$$\Delta\tilde{Q}_1 = \max\left(\sqrt{1-\tilde{\lambda}}\Delta Q, \Delta N_1, \sqrt{\tilde{\lambda}/2}\right)$$

This formula must still be maximized over ΔP , ΔQ , ΔM and ΔN and minimized over ΔM_1 and ΔN_1 . We treat ΔM and ΔN differently from ΔM_1 and ΔN_1 because the quantum capacity is the maximum difference of the log of the effective output dimension minus the log of the effective environment dimension. When evaluating the environment's dimension, we must maximize over ΔM_1 and ΔN_1 , which corresponds to making a measurement choice for the environment, but this becomes a minimization in the overall formula.

As shown in Supplementary Information, when equation (5) is evaluated for the attenuation channel ($N_E = 0$) with unlimited input power, $Q = \log(\lambda) - \log(1 - \lambda)$ for $\lambda \geq 1/2$. This is exactly the quantum capacity of the Gaussian attenuation channel¹⁹. In fact, our model offers slightly more information. Throughout, for our model to make sense the estimates of the number of levels transmitted must be positive integers. The output dimension for this channel is

$$\frac{\lambda\sigma_P\sigma_Q}{\Delta\tilde{P}\Delta\tilde{Q}} = 2\lambda\sigma_P\sigma_Q \geq 1$$

Using $W \geq (\sigma_P^2 + \sigma_Q^2)/2 \geq \sigma_P\sigma_Q$ yields $2\lambda W \geq 2\lambda\sigma_P\sigma_Q \geq 1$, such that $W \geq 1/2\lambda$, which suggests the minimum power necessary to approach capacity.

Our model provides a simple understanding of the recently discovered trade-off between simultaneous classical and quantum communication in optical channels²⁰. Switching between the optimal communication strategies for the two types of information ('time sharing') gives a linear trade-off between the capacities. In ref. 20, it was shown that the rate region, which contains the simultaneously achievable rate pairs for quantum and classical communication, is substantially larger than this naive strategy suggests for an attenuation channel. This can be understood in terms of the DQM for the channel pictured in Fig. 3e. Our model for the attenuation channel with transmissivity $\lambda \geq 1/2$ and signal power W has $C = \log(2\lambda W)$ input bits that get transmitted faithfully to the receiver. The $\log(2(1-\lambda)W)$ highest-order input bits also get transmitted to the environment. These bits cannot transmit quantum information, and so we have $Q = C - \log(2(1-\lambda)W)$. However, they can be used to send classical information, even while transmitting quantum information at full capacity Q , which results in the horizontal portion of the trade-off curve in Fig. 3f. Only after sending classical information at a rate greater than $C - Q$ does a linear trade-off emerge, because the lower-order Q bits can each be used to transmit either classical or quantum information, but not both.

Our approach allows us to solve several vexing questions intractable in the orthodox model. For example, we can exactly solve models for

single-user communication with thermal noise, as well as for multi-user networks including broadcast and multiple-access channels. Analyses in the multi-user setting are typically intractable, even classically^{21,22}. In Supplementary Information, we derive rate regions for the two-input/one-output multiple-access channel and the one-input/two-output broadcast channel (Fig. 4).

We argue that our model also is sophisticated enough to capture the relevant behaviour of quantum communication systems: comparison between the few solvable Gaussian examples and our predictions shows that they always agree to within one or two bits. Furthermore, for examples in which only lower bounds are available, we find good agreement with these, suggesting that known lower bounds are very close to the ultimate capacities. Our rates ought to be achievable in the real channels our models represent. The coding schemes come directly from our optimized input states: each rectangle corresponds to a coherent (possibly squeezed) state, and these states form a basis of a code for the real channel. Small imperfections in the model are eliminated by random coding over the near-perfect typical space arising from many channel uses.

Our model makes some previously known but counterintuitive facts almost obvious. For instance, we can explain why, whereas the classical capacity rises without bound as power increases, the quantum capacity

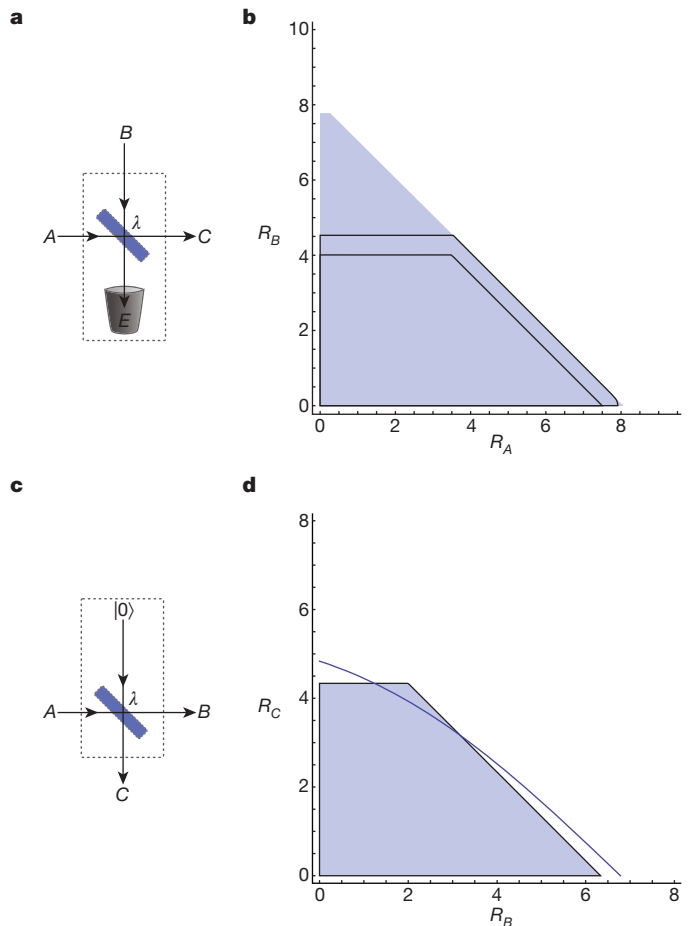


Figure 4 | Multi-user channels. **a**, Two-input/one output multiple-access channel. A and B are the inputs to the channel, C is the output and E is the environment. **b**, Classical-capacity rate region for the multiple-access channel with $\lambda = 0.9$. R_A and R_B are the communication rates to C from inputs A and B, respectively. The outer blue region is the upper bound from ref. 28. The black lines indicate the rate region achievable with coherent states in ref. 28 and the achievable region for our model with $\Delta P_A = \Delta Q_A$ and $\Delta P_B = \Delta Q_B$. Our region is smaller by just a fraction of a bit. **c**, One-input/two-output broadcast channel. **d**, Our calculated rate region is plotted for $W = 50.5$ and $\lambda = 0.8$ (blue region) and compared with a lower bound conjectured to be the actual capacity²⁹.

saturates: increasing power enhances transmission to both receiver and environment in equal measure. Put simply, when trying to communicate privately, shouting your secrets is a poor strategy. We have also explained why time sharing is not optimal for the simultaneous classical–quantum trade-off²⁰.

We also make some predictions. Within our model, entanglement between channel uses and other quantum strategies are not useful. In particular, privacy and coherence are equivalent in our model. We therefore predict that the private and quantum capacities will always be nearly equal in Gaussian channels, even though they can be very different in general²³. Also, two-way communication does not increase the capacity much for Gaussian channels, again counter to the general case²⁴. Furthermore, Gaussian channels should have a simple capacity formula involving a single channel use that is accurate to within a small number of bits (compare with refs 25, 26). Finally, the capacities of Gaussian channels are nearly additive (unlike the extreme superadditivity in refs 25, 26).

The fact that this additivity is not exact can, however, obscure some interesting effects. We know, for example, that Gaussian channels exhibit superactivation; that is, there exist pairs of Gaussian channels each with zero capacity that can nevertheless be used together to achieve positive capacity²⁷. This seems to contradict our prediction of additivity. The resolution is that the resulting capacities are minuscule (the joint rate achieved in ref. 27 is only 0.06 bits). Our predictions are meant to be accurate to within only a few bits, and so there is no contradiction and our results should be asymptotically correct at high signal-to-noise ratio.

Finally, by considering many modes instead of just one as we have above, we recover equation (1). In keeping with our discretization of phase space, the bandwidth, B , can be thought of as the number of modes available, each with noise power $N = \mu^2$. Each can be used individually but requires a proportion of the power budget. By dividing the power equally into B modes and summing these B contributions from equation (4), we get

$$C = B \log \left(\frac{W}{NB} \right)$$

For $W \gg NB$, this expression for C approaches that in equation (1).

Received 8 July; accepted 17 September 2013.

Published online 17 November 2013.

- Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
- Gordon, J. P. in *Quantum Electronics and Coherent Light* (ed. Miles, P. A.) 156 (1964).
- Holevo, A. S. in *Proc. 2nd Jpn–USSR Symp. Probability Theory* (eds Maruyama, G. & Prokhorov, J. V.) 104–119 (Lect. Notes Math. 330, Springer, 1973).
- Schumacher, B. & Westmoreland, M. D. Sending classical information via noisy quantum channels. *Phys. Rev. A* **56**, 131–138 (1997).
- Holevo, A. S. The capacity of a quantum channel with general signal states. *IEEE Trans. Inf. Theory* **44**, 269–273 (1998).

- Holevo, A. S. & Werner, R. F. Evaluating capacities of bosonic gaussian channels. *Phys. Rev. A* **63**, 032312 (2001).
- Richardson, T. & Urbanke, R. The renaissance of Gallager's low-density parity-check codes. *IEEE Commun. Mag.* **41**, 126–131 (2003).
- Devetak, I. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory* **51**, 44–55 (2005).
- Lloyd, S. Capacity of the noisy quantum channel. *Phys. Rev. A* **55**, 1613–1622 (1997).
- Shor, P. W. & Smolin, J. A. Quantum error-correcting codes need not completely reveal the error syndrome. Preprint at <http://arxiv.org/abs/quant-ph/9604006> (1996).
- DiVincenzo, D., Shor, P. W. & Smolin, J. A. Quantum channel capacity of very noisy channels. *Phys. Rev. A* **57**, 830–839 (1998).
- Hastings, M. B. Superadditivity of communication capacity using entangled inputs. *Nature Phys.* **5**, 255–257 (2009).
- Smith, G., Renes, J. M. & Smolin, J. A. Structured codes improve the Bennett–Brassard-84 quantum key rate. *Phys. Rev. Lett.* **100**, 170502 (2008).
- Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).
- Cover, T. M. Comments on broadcast channels. *IEEE Trans. Inf. Theory* **44**, 2524–2530 (1998).
- Avestimehr, S., Diggavi, S. & Tse, D. in *Proc. Allerton Conf. Commun. Control Comput.* 721–728 (IEEE, 2007).
- Eisert, J. & Wolf, M. M. in *Quantum Information with Continuous Variables of Atoms and Light* (eds Cerf, N. J., Leuchs, G. & Polzik, E. S.) 23–42 (Imperial Coll. Press, 2007).
- Giovannetti, V. et al. Classical capacity of the lossy bosonic channel: the exact solution. *Phys. Rev. Lett.* **92**, 027902 (2004).
- Wolf, M. M., Perez-Garcia, D. & Giedke, G. Quantum capacities of bosonic channels. *Phys. Rev. Lett.* **98**, 130501 (2007).
- Wilde, M. M., Hayden, P. & Guha, S. Information trade-offs for optical quantum communication. *Phys. Rev. Lett.* **108**, 140501 (2012).
- Cover, T. M. & Thomas, J. A. *Elements of Information Theory* 374–458 (Wiley, 1991).
- El Gamal, A. & Kim, Y.-H. *Network Information Theory* (Cambridge Univ. Press, 2011).
- Horodecki, K., Horodecki, M., Horodecki, P. & Oppenheim, J. Secure key from bound entanglement. *Phys. Rev. Lett.* **94**, 160502 (2005).
- Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
- Smith, G. & Yard, J. Quantum communication with zero-capacity channels. *Science* **351**, 1812–1815 (2008).
- Smith, G. & Smolin, J. A. Extensive nonadditivity of privacy. *Phys. Rev. Lett.* **103**, 120503 (2009).
- Smith, G., Smolin, J. A. & Yard, J. Gaussian bosonic synergy: quantum communication via realistic channels of zero quantum capacity. *Nature Photon.* **5**, 624–627 (2011).
- Yen, B. J. & Shapiro, J. H. Multiple-access bosonic communications. *Phys. Rev. A* **72**, 062312 (2005).
- Guha, S., Shapiro, J. H. & Erkmen, B. I. Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture. *Phys. Rev. A* **76**, 032303 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the DARPA QUEST programme under contract no. HR0011-09-C-0047.

Author Contributions G.S. and J.A.S. designed the research and carried out the research and computations. Both authors wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S. (gsbsmith@gmail.com).

Increased insolation threshold for runaway greenhouse processes on Earth-like planets

Jérémy Leconte¹, François Forget¹, Benjamin Charnay¹, Robin Wordsworth² & Alizée Pottier¹

The increase in solar luminosity over geological timescales should warm the Earth's climate, increasing water evaporation, which will in turn enhance the atmospheric greenhouse effect. Above a certain critical insolation, this destabilizing greenhouse feedback can 'run away' until the oceans have completely evaporated^{1–4}. Through increases in stratospheric humidity, warming may also cause evaporative loss of the oceans to space before the runaway greenhouse state occurs^{5,6}. The critical insolation thresholds for these processes, however, remain uncertain because they have so far been evaluated using one-dimensional models that cannot account for the dynamical and cloud feedback effects that are key stabilizing features of the Earth's climate. Here we use a three-dimensional global climate model to show that the insolation threshold for the runaway greenhouse state to occur is about 375 W m^{-2} , which is significantly higher than previously thought^{6,7}. Our model is specifically developed to quantify the climate response of Earth-like planets to increased insolation in hot and extremely moist atmospheres. In contrast with previous studies, we find that clouds have a destabilizing feedback effect on the long-term warming. However, subsident, unsaturated regions created by the Hadley circulation have a stabilizing effect that is strong enough to shift the runaway greenhouse limit to higher values of insolation than are inferred from one-dimensional models. Furthermore, because of wavelength-dependent radiative effects, the stratosphere remains sufficiently cold and dry to hamper the

escape of atmospheric water, even at large fluxes. This has strong implications for the possibility of liquid water existing on Venus early in its history, and extends the size of the habitable zone around other stars.

Planetary atmospheres naturally settle into a thermal equilibrium state where their outgoing thermal emission balances the absorbed part of the incoming sunlight. The resulting climate is stabilized by the fact that a temperature increase results in enhanced cooling by means of thermal emission. When a condensable greenhouse gas is present at the planet's surface, as is the case for water on the Earth, this stabilizing feedback is hampered by the destabilizing greenhouse feedback: evaporation and, thus, the water-vapour greenhouse effect increase with temperature, reducing the cooling. Under present Earth conditions, this greenhouse feedback is both strong enough to maintain clement surface temperatures and weak enough for the climate to remain stable.

When solar heating becomes stronger, however, water vapour can become abundant enough to make the atmosphere optically thick at all thermal wavelengths^{7,8}. Thermal flux then originates in the upper troposphere only and reaches a maximum, $\sim 282 \text{ W m}^{-2}$, that is independent of the surface temperature⁹. If the planet absorbs more than this critical flux, thermal equilibrium can be restored only by vaporizing all the water available and reaching high surface temperatures at which the surface starts to radiate at visible wavelengths^{4,7}. This is the runaway greenhouse state^{1–4}.

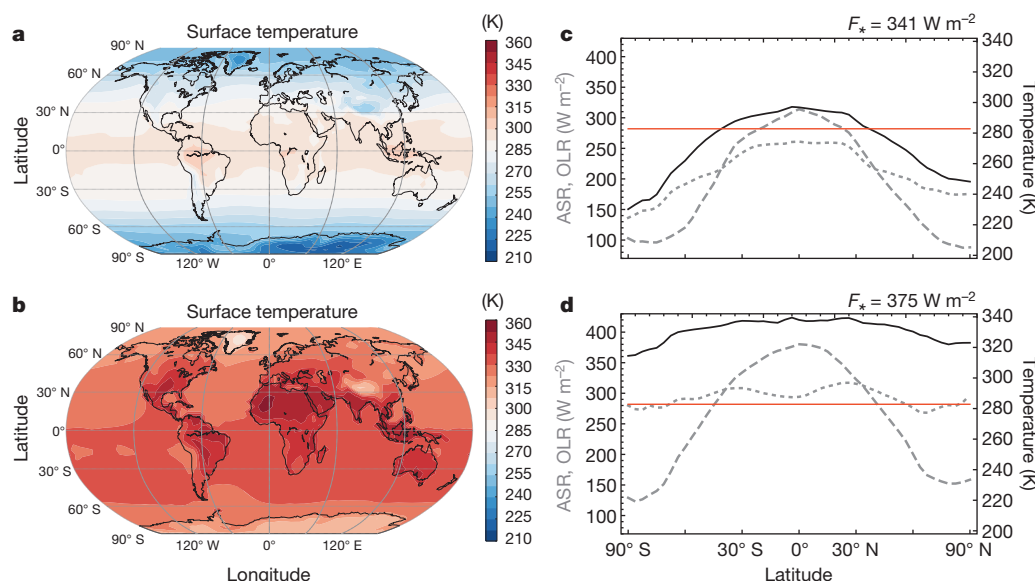


Figure 1 | Temperature and radiative budget for the Earth under two insolation levels. **a, b**, Maps of the annual mean surface temperature for the models corresponding to present Earth ($F_{\star} = 341 \text{ W m}^{-2}$; **a**) and to a mean solar flux of 375 W m^{-2} (**b**), just before the runaway greenhouse instability is triggered. **c, d**, Zonally and annually averaged surface temperature (solid black), absorbed

stellar radiation (ASR; grey dashed) and outgoing long-wave radiation (OLR; grey dotted) for the insolation levels in **a** (**c**) and **b** (**d**). The red horizontal line shows the radiation limit on the emitted flux for a saturated water atmosphere (282 W m^{-2}). As visible in the hot case, unsaturated subtropical regions allow the atmosphere to emit more than this limit.

¹Laboratoire de Météorologie Dynamique, Institut Pierre Simon Laplace, 4 Place Jussieu, BP 99, 75252 Paris, France. ²Department of Geological Sciences, University of Chicago, Chicago, Illinois 60637, USA.

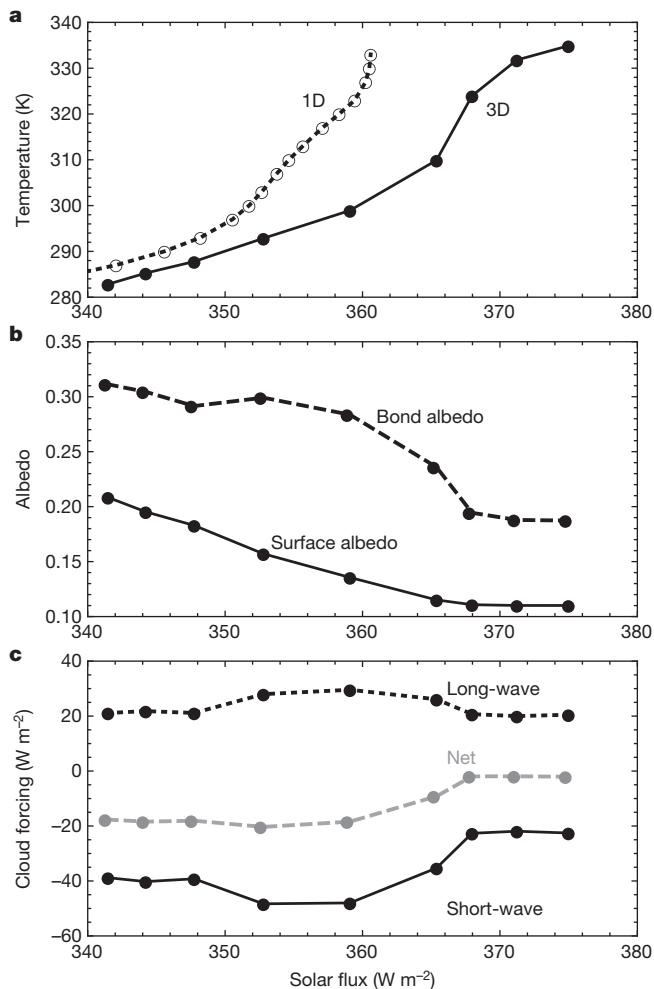


Figure 2 | Evolution of the mean surface temperature, planetary albedo and cloud radiative forcing with the mean solar incoming flux. Curves start from present Earth conditions ($\sim 341 \text{ W m}^{-2}$). **a**, Average surface temperature in our 3D baseline model (solid) and temperature in the 1D cloud-free model (dashed). **b**, Average surface albedo (solid) and effective planetary albedo (or Bond albedo; dashed). **c**, Short-wave (solid), long-wave (dotted) and net (grey dashed) cloud radiative forcing in the baseline model.

Because it has mostly been studied using one-dimensional (1D) atmosphere models, the aforementioned mechanism strongly relies on the assumption that the troposphere is saturated in water vapour. Furthermore, by construction, these studies could not account for spatial inhomogeneities in insolation and in the resulting water vapour and cloud distributions^{4,6,7}. To overcome these limitations, we have developed a three-dimensional (3D) global climate model fit to describe hot atmospheres in which water vapour can become a dominant species^{10–13} (Methods). The main challenges of such a model are threefold. First, the radiative transfer must be fast, yet able to describe accurately the spectroscopic properties of various gases in a wide temperature–pressure domain. Second, the modelling of the physical processes that are not specific to a given planet (convection, turbulence and so on) must rely on the fewest free parameters possible to ensure its validity under stringent conditions. Finally, for hot, moist atmospheres, the description of the water cycle (moist convection, cloud formation and so on) must take into account the fact that water can become a major constituent of the atmosphere. For these reasons, the global climate models usually used to predict Earth climate are generally not suited for such studies.

Here we perform simulations of future Earth climate by running our baseline model for various (increasing) values of the solar constant until radiative balance is achieved. For the present solar flux ($F_{\star} \approx 341 \text{ W m}^{-2}$), our generic model reproduces the energy budget and the characteristics of our climate¹³ (Fig. 1). When the flux is increased, the planet undergoes a decrease in surface albedo that is due to the melting of the permanent polar ice caps and the reduced seasonal snow cover (Fig. 2). For fluxes greater than $\sim 350 \text{ W m}^{-2}$, only seasonal ice caps appear during the polar night. The amount of water vapour also increases. This results in a more efficient absorption of the incoming stellar light, but also in an enhanced greenhouse effect, which tends to homogenize the surface temperatures. Although continental surfaces can reach temperatures of around 100°C because of the intense solar and greenhouse heating, sea surface temperatures remain moderate with a small diurnal variation because they are thermodynamically controlled by latent heat cooling¹⁴. Finally, for fluxes greater than $\sim 375 \text{ W m}^{-2}$, no thermal equilibrium exists. Although surface temperature increases with time, thermal emission reaches a limiting value. This is the onset of the runaway greenhouse instability.

This runaway greenhouse insolation threshold is greater than those recently found by previous 1D studies^{6,7} and confirmed by our 1D model (Extended Data Fig. 1). To understand the mechanisms that increase this threshold in practice, we first analyse the radiative effect of clouds. Although 1D simulations cannot properly capture spatial variations in cloud distribution, it has been suggested that clouds should have a feedback effect that stabilizes the climate against the runaway greenhouse

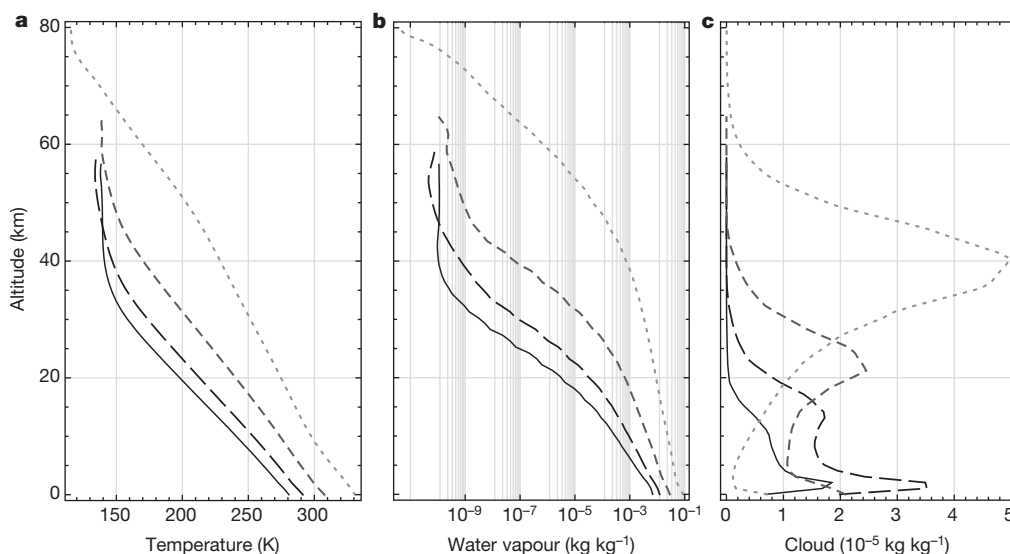


Figure 3 | Evolution of globally averaged vertical profiles. Mean vertical profiles of temperature (**a**) and water-vapour (**b**) and condensed-water (**c**) mixing ratios (in kilograms per kilogram of moist air in **b** and **c**) for four different insulations: 341 W m^{-2} (solid); 353 W m^{-2} (long dashed); 365 W m^{-2} (dashed); 375 W m^{-2} (dotted).

effect^{4–6}. This tentative conclusion was based on the fact that the present net cloud radiative forcing on the Earth is negative, meaning that the albedo increase due to low-level clouds exceeds the greenhouse effect of high-level clouds¹⁵. Because of the increased evaporation resulting from the warming, cloud thickness might therefore increase and enhance the stabilizing effect of clouds^{4–6}.

Our simulations of the evolution of radiative cloud forcing with insolation suggest the opposite (Fig. 2c). This is due to a displacement of the cloud formation region towards higher altitudes (Fig. 3). As a result, the temperature at the mean cloud emission level increases much less with insolation than does the surface temperature. Even though the cloud optical depth increases, the greenhouse feedback of the clouds exceeds their albedo effect. At higher fluxes, clouds become thinner, which reduces both long-wave and short-wave radiative cloud forcing. For the reason described above, however, the greenhouse effect of clouds prevails and the net radiative forcing tends to vanish.

There are several reasons for the vertical displacement of clouds. First, moist convection and Hadley circulation become more intense as the insolation increases, extending the troposphere. Second, to form and persist, clouds need to be able to lose the latent heat released during condensation. Because of the infrared opacity increase of the atmosphere, the altitude at which clouds can efficiently cool radiatively rises. This may explain both the progressive disappearance of low-level clouds and the small change in cloud-deck temperature seen in Fig. 3.

Although the tendencies described above should be robust, the precise value of the cloud radiative forcing does depend on the assumptions made about the cloud microphysics. In our baseline model, for instance, we assume that the number density of cloud condensation nuclei, that is, the number of cloud particles per unit mass of air, remains fixed at a value that is representative of the modern Earth. As is expected, a larger mass mixing ratio of condensed water thus entails bigger cloud particles, which precipitate more easily and have a smaller radiative effect. To explore, and put limits on, the possible behaviour of clouds, we conducted a set of simulations in which the radii of cloud particles were kept constant. This assumption results in smaller radii, and thus overestimates cloud optical depths and both short- and long-wave cloud forcing. As can be seen in Extended Data Fig. 2, long-wave and short-wave forcing indeed increase continuously with insolation. However, for the reasons mentioned above, the greenhouse effect of clouds eventually overcomes the albedo effect. Therefore, the fact that the cloud feedback has a destabilizing effect under extreme insolation seems a robust conclusion, and one that supports results obtained in the context of anthropogenic global warming¹⁶.

Clouds might not help stabilize climate against solar flux increases, but atmospheric dynamics does. This is due to the fact that when a parcel of moist air is heated without any source of moisture, its water vapour pressure decreases relative to the saturation pressure. As a result, the Earth's troposphere itself is not saturated everywhere, unlike what is often assumed in 1D models. An example of this are sub-tropical regions receiving hot air that comes from the Hadley cell, which dries during its ascent in the tropics and is compressed adiabatically during its descent¹⁷ (Fig. 4). Because they stabilize the Earth's tropics today, such dynamically unsaturated regions where the water-vapour greenhouse effect is reduced stabilize the climate against the runaway greenhouse effect by playing the role of 'radiative fins' where the emission can locally exceed the maximum emission for a saturated atmosphere^{18,19} (Fig. 1).

Furthermore, these unsaturated regions have the interesting property of extending upwards and polewards with the Hadley cell (Fig. 4), and of getting dryer when the insolation increases¹⁷ (Extended Data Figs 3 and 4). Such unsaturated regions, which cannot be predicted using a 1D model, ensure that the infrared photosphere is always at a lower altitude than in the saturated case, yielding a more efficient cooling to space. This stabilizing feedback is the main reason why the runaway greenhouse insolation threshold predicted by 3D simulations is greater than previously found. And it would be even greater if it were not for the positive cloud feedback (Methods).

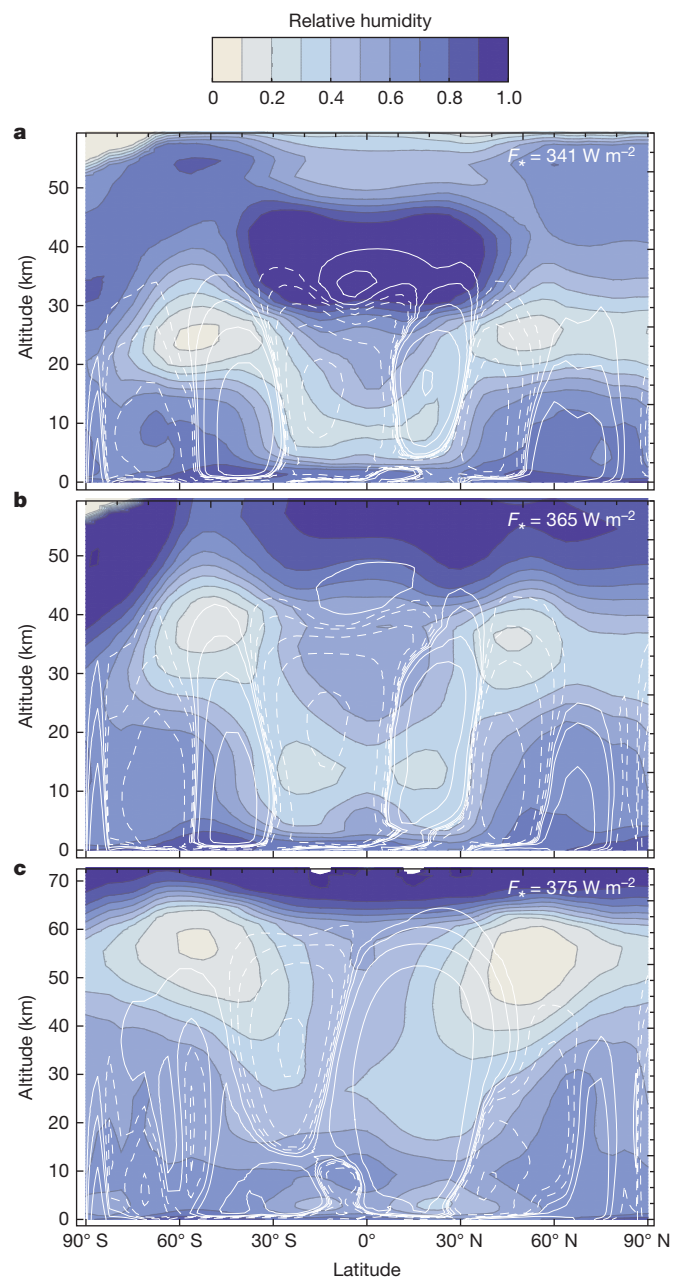


Figure 4 | Meridional distribution of relative humidity. Zonally and annually averaged relative humidity for models in which the respective stellar fluxes are 341 (a), 365 (b) and 375 (c) W m^{-2} . White contours are streamlines showing the Hadley circulation (solid and dashed curves depict clockwise and anticlockwise rotation, respectively).

With this new estimation, the inner edge of the habitable zone for Earth-like planets in the Solar System is pushed inwards to ~ 0.95 AU, which means that the Earth should not enter a runaway greenhouse state for at least another billion years²⁰. However, the question of whether or not the warming-induced increase in the stratospheric temperature and humidity could cause the loss of the oceans to atmospheric escape before the runaway greenhouse instability is triggered, as suggested by some 1D studies^{4,6}, is still subject to debate. Our simulations tend to answer this question in the negative. This is due to both non-grey radiative effects^{14,21,22} (Methods) and unsaturated regions that flatten the thermal profile in the troposphere. As a result, stratospheric temperatures are much lower than anticipated and stratospheric humidity cannot reach the threshold needed for efficient water photodissociation and escape of hydrogen to space (Fig. 3).

As well as extending the habitable zone²³ towards the Sun, our results, more notably, highlight the fact that global climate models are needed to understand subtle climate feedbacks resulting from the inhomogeneous insolation of planetary surfaces that are not amenable to 1D modelling. Although adding complexity and uncertainties, these subtle effects must be accounted for when modelling real planets, especially tidally synchronized exoplanets where they are even more pronounced^{12,24,25}. In particular, although our simulation suggest that Venus, if it had the Earth's rotation rate, would have been in a runaway greenhouse state since its formation, its slow retrograde motion (because it results in a long Solar day) or a small water inventory^{12,26} could completely change that view: both of these effects would tend to stabilize planetary climates against a runaway greenhouse state, respectively by increasing the albedo effect of clouds or by reducing the greenhouse effect of water vapour.

METHODS SUMMARY

Our simulations were performed with an upgraded version of the LMD generic global climate model specifically developed for the study of extrasolar planets^{10,12,27} and palaeoclimates^{11,28}. The model uses the 3D dynamical core of the LMDZ Earth global climate model used in studies by the Intergovernmental Panel on Climate Change studies²⁹, which is based on a finite-difference formulation of the primitive equations of geophysical fluid dynamics.

General physical processes relevant for present-day Earth—including ground thermal inertia and albedo (Extended Data Fig. 5), turbulent transport, dry convection, evaporation/condensation and precipitation—are parameterized in the most physically based way to ensure the robustness of the model under extreme conditions¹³. For the present study, special care has been taken to treat properly the situation in which water vapour can become a major constituent of the atmosphere. As detailed in Methods, the radiative transfer and moist convection scheme have been validated by recent 1D results for the hot, pure-water atmosphere regime (Extended Data Fig. 1). A numerical scheme accounting for change in atmospheric mass and surface pressure with water-vapour evaporation/condensation has also been implemented.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 August; accepted 4 November 2013.

- Simpson, G. C. Some studies in terrestrial radiation. *Mem. R. Meteorol. Soc.* **11**, 69–95 (1927).
- Komabayashi, M. Discrete equilibrium temperatures of a hypothetical planet with the atmosphere and the hydrosphere of one component-two phase system under constant solar radiation. *J. Meteorol. Soc. Jpn* **45**, 137–138 (1967).
- Ingersoll, A. P. The runaway greenhouse: a history of water on Venus. *J. Atmos. Sci.* **26**, 1191–1198 (1969).
- Kasting, J. F., Pollack, J. B. & Ackerman, T. P. Response of Earth's atmosphere to increases in solar flux and implications for loss of water from Venus. *Icarus* **57**, 335–355 (1984).
- Kasting, J. F. Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus* **74**, 472–494 (1988).
- Kopparapu, R. K. *et al.* Habitable zones around main-sequence stars: new estimates. *Astrophys. J.* **765**, 131 (2013).
- Goldblatt, C., Robinson, T. D., Zahnle, K. J. & Crisp, D. Low simulated radiation limit for runaway greenhouse climates. *Nature Geosci.* **6**, 661–667 (2013).
- Goldblatt, C. & Watson, A. J. The runaway greenhouse: implications for future climate change, geoengineering and planetary atmospheres. *Phil. Trans. R. Soc. Lond. A* **370**, 4197–4216 (2012).
- Nakajima, S., Hayashi, Y.-Y. & Abe, Y. A study on the 'runaway greenhouse effect' with a one-dimensional radiative-convective equilibrium model. *J. Atmos. Sci.* **49**, 2256–2266 (1992).
- Wordsworth, R. D. *et al.* Gliese 581d is the first discovered terrestrial-mass exoplanet in the habitable zone. *Astrophys. J.* **733**, L48 (2011).
- Forget, F. *et al.* 3D modelling of the early Martian climate under a denser CO₂ atmosphere: temperatures and CO₂ ice clouds. *Icarus* **222**, 81–99 (2013).
- Leconte, J. *et al.* 3D climate modeling of close-in land planets: circulation patterns, climate moist bistability, and habitability. *Astron. Astrophys.* **554**, A69 (2013).
- Charnay, B. *et al.* Exploring the faint young Sun problem and the possible climates of the Archean Earth with a 3D GCM. *J. Geophys. Res.* **118**, 414–431 (2013).
- Pierrehumbert, R. T. *Principles of Planetary Climate* 289–299 (Cambridge Univ. Press, 2010).
- Ramanathan, V. *et al.* Cloud-radiative forcing and climate: results from the Earth radiation budget experiment. *Science* **243**, 57–63 (1989).
- Soden, B. J. & Held, I. M. An assessment of climate feedbacks in coupled ocean atmosphere models. *J. Clim.* **19**, 3354–3360 (2006).
- Sherwood, S. C. *et al.* Relative humidity changes in a warmer climate. *J. Geophys. Res.* **115**, D09104 (2010).
- Pierrehumbert, R. T. Thermostats, radiator fins, and the local runaway greenhouse. *J. Atmos. Sci.* **52**, 1784–1806 (1995).
- Ishiwatari, M., Takehiro, S.-I., Nakajima, K. & Hayashi, Y.-Y. A numerical study on appearance of the runaway greenhouse state of a three-dimensional gray atmosphere. *J. Atmos. Sci.* **59**, 3223–3238 (2002).
- Gough, D. O. Solar interior structure and luminosity variations. *Sol. Phys.* **74**, 21–34 (1981).
- Arking, A. & Grossman, K. The influence of line shape and band structure on temperatures in planetary atmospheres. *J. Atmos. Sci.* **29**, 937–949 (1972).
- Wordsworth, R. & Pierrehumbert, R. T. Water loss from terrestrial planets with CO₂-rich atmospheres. *Astrophys. J.* **778**, 154 (2013).
- Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. Habitable zones around main sequence stars. *Icarus* **101**, 108–128 (1993).
- Pierrehumbert, R. T. A palette of climates for Gliese 581g. *Astrophys. J.* **726**, L8 (2011).
- Yang, J., Cowan, N. B. & Abbot, D. S. Stabilizing cloud feedback dramatically expands the habitable zone of tidally locked planets. *Astrophys. J.* **771**, L45 (2013).
- Abe, Y., Abe-Ouchi, A., Sleep, N. H. & Zahnle, K. J. Habitable zone limits for dry planets. *Astrobiology* **11**, 443–460 (2011).
- Selsis, F., Wordsworth, R. D. & Forget, F. Thermal phase curves of nontransiting terrestrial exoplanets. I. Characterizing atmospheres. *Astron. Astrophys.* **532**, A1 (2011).
- Wordsworth, R. *et al.* Global modelling of the early Martian climate under a denser CO₂ atmosphere: water cycle and ice evolution. *Icarus* **222**, 1–19 (2013).
- Hourdin, F. *et al.* The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Clim. Dyn.* **27**, 787–813 (2006).

Acknowledgements We thank our referees, J. Kasting and Y. Abe, for their thorough review, and A. Spiga and F. Selsis for discussions. This work was supported by grants from Région Ile-de-France.

Author Contributions J.L. developed the 'high temperature/humidity' version of the generic global climate model, performed the calculations, and led the analysis and writing of the results. F.F. initiated the development of the generic global climate model and provided critical advice during analysis and writing. B.C. worked on the development of the model and helped perform the comparison with present Earth climatology. R.W. developed the original version of the generic model and implemented the radiative transfer scheme. A.P. performed comparison runs and sensitivity studies. All the authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L. (jeremy.leconte@lmd.jussieu.fr).

Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning

Flavio Donato¹, Santiago Belluco Rompani¹ & Pico Caroni¹

Learning and memory processes can be influenced by recent experience, but the mechanisms involved are poorly understood. Enhanced plasticity during critical periods of early life is linked to differentiating parvalbumin (PV)-interneuron networks^{1–7}, suggesting that recent experience may modulate learning by targeting the differentiation state of PV neurons^{8–11} in the adult. Here we show that environmental enrichment and Pavlovian contextual fear conditioning induce opposite, sustained and reversible hippocampal PV-network configurations in adult mice. Specifically, enrichment promotes the emergence of large fractions of low-differentiation (low PV and GAD67 expression) basket cells with low excitatory-to-inhibitory synaptic-density ratios, whereas fear conditioning leads to large fractions of high-differentiation (high PV and GAD67 expression) basket cells with high excitatory-to-inhibitory synaptic-density ratios. Pharmacogenetic inhibition or activation of PV neurons was sufficient to induce such opposite low-PV-network or high-PV-network configurations, respectively. The low-PV-network configuration enhanced structural synaptic plasticity^{12,13}, and memory consolidation and retrieval, whereas these were reduced by the high-PV-network configuration. We then show that maze navigation learning¹⁴ induces a hippocampal low-PV-network configuration paralleled by enhanced memory and structural synaptic plasticity throughout training, followed by a shift to a high-PV-network configuration after learning completion. The shift to a low-PV-network configuration specifically involved increased vasoactive intestinal polypeptide (VIP)-positive GABAergic boutons and synaptic transmission onto PV neurons^{15,16}. Closely comparable low- and high-PV-network configurations involving VIP boutons were specifically induced in primary motor cortex upon rotarod motor learning^{17,18}. These results uncover a network plasticity mechanism induced after learning through VIP-PV microcircuit modulation¹⁹, and involving large, sustained and reversible shifts in the configuration of PV basket-cell networks in the adult. This novel form of experience-related plasticity in the adult modulates memory consolidation, retrieval and learning, and might be harnessed for therapeutic strategies to promote cognitive enhancement and neuroprotection.

To determine whether plasticity and learning in the adult may involve alterations in the differentiation state of PV-positive GABA (γ -aminobutyric acid)-ergic interneurons, we monitored PV immunoreactivity in ensembles of genetically identified PV neurons near the pyramidal cell layer in dorsal hippocampal area CA3b (Supplementary Discussion). We initially compared cage control to environmentally enriched, and to contextually fear conditioned mice. The environmental enrichment protocol enhances hippocampal plasticity and learning^{12,13}, whereas fear specifically restricted to training context depends on enhanced filopodial synapses onto PV neurons by hippocampal mossy fibres upon contextual fear conditioning²⁰. The PV-neuron group with lowest PV signals (low-PV) represented $2 \pm 0.8\%$ of the total number of PV neurons, whereas the group with highest PV levels (high-PV) represented $19 \pm 2\%$ of the total number of PV neurons ($n = 15$ mice). Housing mice under environmentally enriched conditions induced a robust increase in low-PV neurons, whereas contextual fear conditioning greatly

enhanced high-PV-neuron contents (Fig. 1a and Supplementary discussion). Total numbers of PV neurons were not affected by these experimental conditions (CA3b PV-neuron numbers per $50,000 \mu\text{m}^2$ within $150 \mu\text{m}$ of the pyramidal cell layer were 20.0 ± 0.9 (control), 19.3 ± 1.4 (environmentally enriched) and 21.2 ± 1.7 (contextual fear conditioning); $n = 18$ mice each). Under all experimental conditions, PV immunoreactivity in individual interneurons was closely correlated to GAD67 immunoreactivity (Pearson correlation, 0.92; Fig. 1b and Extended Data Fig. 2). GAD67 is a key GABA-synthesizing enzyme that links synaptic activity to GABA levels, regulates PV neuron synaptogenesis and is a differentiation marker of PV neurons²¹. An analysis of PV neurons in postnatal hippocampal CA3b revealed that at the time of eye opening, PV distributions resemble the low-PV configuration induced in the adult after environmental enrichment (Fig. 1c and Extended Data Fig. 3).

To investigate how defined subpopulations of PV neurons are influenced by experience, we analysed presynaptic terminals of basket and chandelier cells, the two main subpopulations of hippocampal PV neurons in CA3. PV immunoreactivity signals were comparable at soma, axonal boutons and dendrites of individual PV basket cells (Fig. 1d). Perisomatic PV-bouton immunoreactivity distributions for individual baskets were closely comparable to those of the whole PV network for each experimental condition (Figs 1e and Extended Data Fig. 4a, b). This was in contrast to the distribution of PV boutons onto axon initial segments established by chandelier cells, which were not detectably affected by experience (Fig. 1f and Extended Data Fig. 4c). Therefore, experience specifically affects PV and GAD67 levels in basket cells.

To investigate behavioural and structural counterparts of PV-network configurations, we probed incidental learning and memory using a familiar-object-recognition protocol¹² (FOR protocol, also known as novel-object recognition; see Supplementary Discussion), and mossy-fibre synapse turnover using a transient protein synthesis inhibition assay *in situ*¹². Environmentally enriched mice exhibited enhanced FOR (Fig. 1g), more c-Fos-positive neurons in CA3b after FOR (Extended Data Fig. 5), and enhanced active-zone turnover at mossy-fibre terminals (Fig. 1g). By contrast, mice that underwent contextual fear conditioning 2 h after acquisition performed worse than controls on FOR retrieval, exhibited less c-Fos-positive neurons in CA3b upon FOR, and exhibited much reduced synapse turnover (Fig. 1g and Extended Data Fig. 5). To rule out stress-related effects on incidental learning and memory, we fear-conditioned mice in the dark; that is, under conditions not involving the hippocampus²². Mice that were fear conditioned in the dark froze in response to a conditioned stimulus (tone), but exhibited no PV network configuration shift in CA3b (Extended Data Fig. 5), and were indistinguishable from controls in FOR (Fig. 1g and Extended Data Fig. 5).

To investigate the mechanisms underlying shifts in PV-network configuration upon experience, we analysed densities of Bassoon and PSD95 excitatory synaptic puncta and of VGAT and gephyrin inhibitory synaptic puncta along PV-positive dendrites in CA3b in PV-Cre \times Rosa-CAG-STOP-tdTomato mice, in which PV neurons were labelled genetically (Fig. 2a and Extended Data Fig. 6a). Fear conditioning nearly

¹Friedrich Miescher Institut, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.

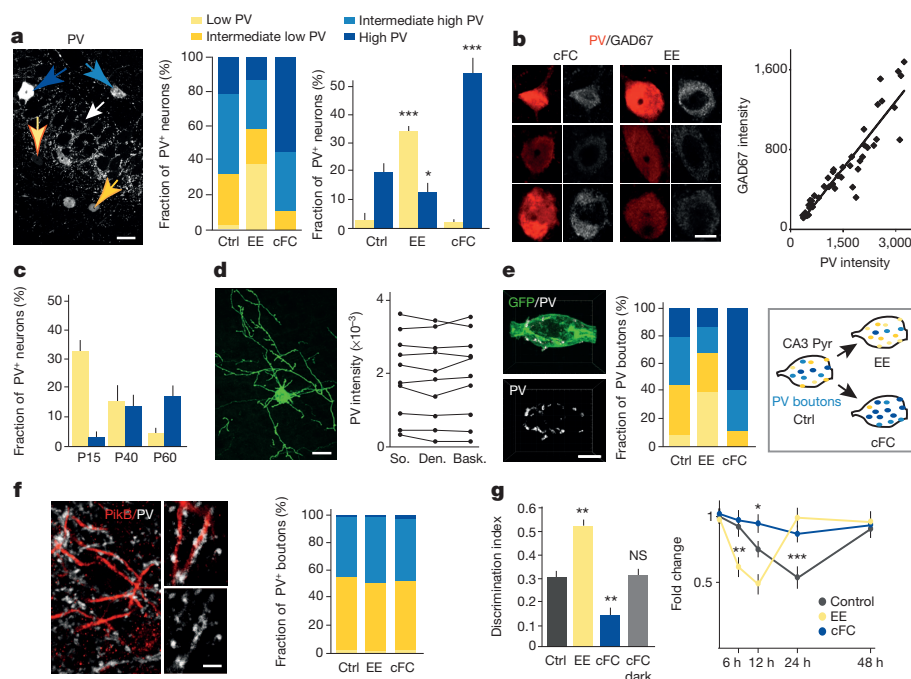


Figure 1 | Differentiation plasticity of hippocampal parvalbumin-expressing basket-cell networks in adult mice. **a**, Experience-related plasticity of PV neurons. Arrows, high-PV-signal (dark blue), intermediate high-PV-signal (blue), intermediate low-PV-signal (gold) and low-PV-signal (yellow) neurons. White arrows indicate PV basket-cell terminals. Relative contents of PV neurons in control (Ctrl), environmentally enriched (EE, 3 weeks) and contextually fear conditioned (cFC, day1) mice. $n = 15$ mice each (50 neurons per mouse). **b**, Differentiation plasticity of PV neurons in the adult. PV- and GAD67-signal relationship in individual PV neurons from Ctrl, EE and cFC mice. $n = 5$ mice each (50 neurons per mouse). **c**, PV-network configuration during maturation of hippocampal CA3 (CA3b). $n = 8$ mice each. Colours as in **a**. **d**, PV immunoreactivity in Soma (So), dendrites (Den) and terminal boutons (Bask) of individual PV-expressing basket cells. $n = 8$ –10 dendrite stretches or 30–50 boutons per cell. **e**, PV-bouton distribution at individual pyramidal neuron perisomatic baskets. $n = 60$ neurons (3 mice) each. Colours as in **a**. **f**, PV-bouton labelling intensities at axon initial segments (chandelier cells). $n = 60$ AIS (3 mice) each. Colours as in **a**. **g**, PV-network configuration correlated to memory consolidation and retrieval (left, FOR), and synapse turnover (right). $n = 8$ mice cFC, and 3 mice and 40 mossy-fibre terminals each (mossy-fibre-terminal active-zone densities). Colours as in **a**. NS, not significant. Values are means \pm s.e.m. Scale bars, 20 μ m (**a**, **d**); 10 μ m (**b**); 5 μ m (**e**, **f**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

doubled average excitatory puncta densities in all strata (Fig. 2b). By contrast, environmental enrichment did not affect excitatory puncta densities (Fig. 2b), but doubled average inhibitory puncta densities onto PV dendrites (Fig. 2b). The marked increase in inhibitory puncta densities upon environmental enrichment affected VIP-positive and VIP-negative^{15,16,19} boutons to comparable extents (Extended Data Fig. 6b).

To investigate whether the relationship between PV-network configuration and excitatory–inhibitory synaptic puncta densities might apply at the level of individual PV neurons, we plotted synaptic puncta densities against PV-labelling intensities for individual PV-dendrite stretches for each experimental condition. Regardless of whether they were plotted within experimental conditions or in pooled form, the data points fitted comparable and robustly significant reciprocal power laws (Pearson correlation Bassoon–PV, 0.90, $P < 0.0001$; Pearson correlation gephyrin–PV, 0.81, $P < 0.0001$; Fig. 2c and Extended Data Fig. 6c). Therefore, excitatory–inhibitory synaptic puncta densities adjust at the level of individual PV neurons, in parallel with alterations in PV- and GAD67-expression levels in those same PV neurons.

To determine whether enhanced excitation of PV neurons might be sufficient to induce a high-PV-network configuration shift we took an optogenetic approach²³ in slice cultures from PV-Cre mice. Direct photostimulation of PV neurons specifically induced a transient high-PV-network configuration 6h after light stimulation (Fig. 2d, e and Extended Data Fig. 7a). To determine whether PV-network configuration shifts are sufficient to influence hippocampal memory and synapse turnover, we took a pharmacogenetic approach *in vivo* (Supplementary Discussion)²⁴. Expressed channels were activated transiently (1.5–2 h (ref. 24)) using a specific ligand delivered systemically at 12-h intervals (Fig. 2f). Ligand delivery and behavioural protocols were scheduled so that ligand-mediated channel activation did not overlap in time with behavioural testing. PV-neuron inhibition was sufficient to specifically induce a hippocampal low-PV-network configuration, enhanced FOR performance and enhanced synapse turnover (Fig. 2f). Similarly, PV-neuron activation was sufficient to specifically induce a hippocampal

high-PV-network configuration, diminished FOR performance and greatly reduced synapse turnover (Fig. 2f, Extended Data Figs 7c, 8).

To determine whether experience-related alterations in hippocampal memory and synaptic plasticity depend on specific hippocampal shifts in PV-network configuration, we exposed mice to environmental enrichment or contextual fear conditioning to induce corresponding PV-network configuration shifts, then induced opposite PV-network shifts through pharmacogenetics. We determined whether this was sufficient to reverse the effects of experience on behaviour and plasticity. Enhancing PV-neuron activation in enriched mice switched their PV neurons to a high-PV-network configuration, and reversed FOR performance to levels comparable to those of fear-conditioned mice. Similarly, inhibiting PV neurons in fear-conditioned mice switched their PV neurons to a low-PV-network configuration, which was sufficient to reverse FOR performance to resemble that of environmentally enriched mice (Fig. 2g). Taken together, these results provided causal evidence that the changes in memory and structural synaptic plasticity induced by experience are due to corresponding alterations in PV-network configuration (Fig. 2h and Supplementary Discussion).

Given that the shift to low-PV-network configuration in CA3 facilitates memory consolidation and retrieval, and that environmental enrichment improves several forms of learning and memory, we reasoned that incremental trial-and-error learning tasks during which animals must effectively combine and retrieve previous task-related memories to adjust and improve performance might involve similar shifts to a low-PV-network configuration. Accordingly, we analysed the composition of PV networks in hippocampal CA3 as mice learned to navigate a Morris water maze (MWM¹⁴). The learning phase of the MWM (Fig. 3a) was accompanied by a shift to a low-PV-network configuration closely comparable to environmental enrichment, whereas the establishment of a robust reference memory at the end of the learning protocol was accompanied by a shift to a high-PV-network closely comparable to contextual fear conditioning (Fig. 3b). Similarly, mice undergoing MWM learning exhibited first higher and then lower

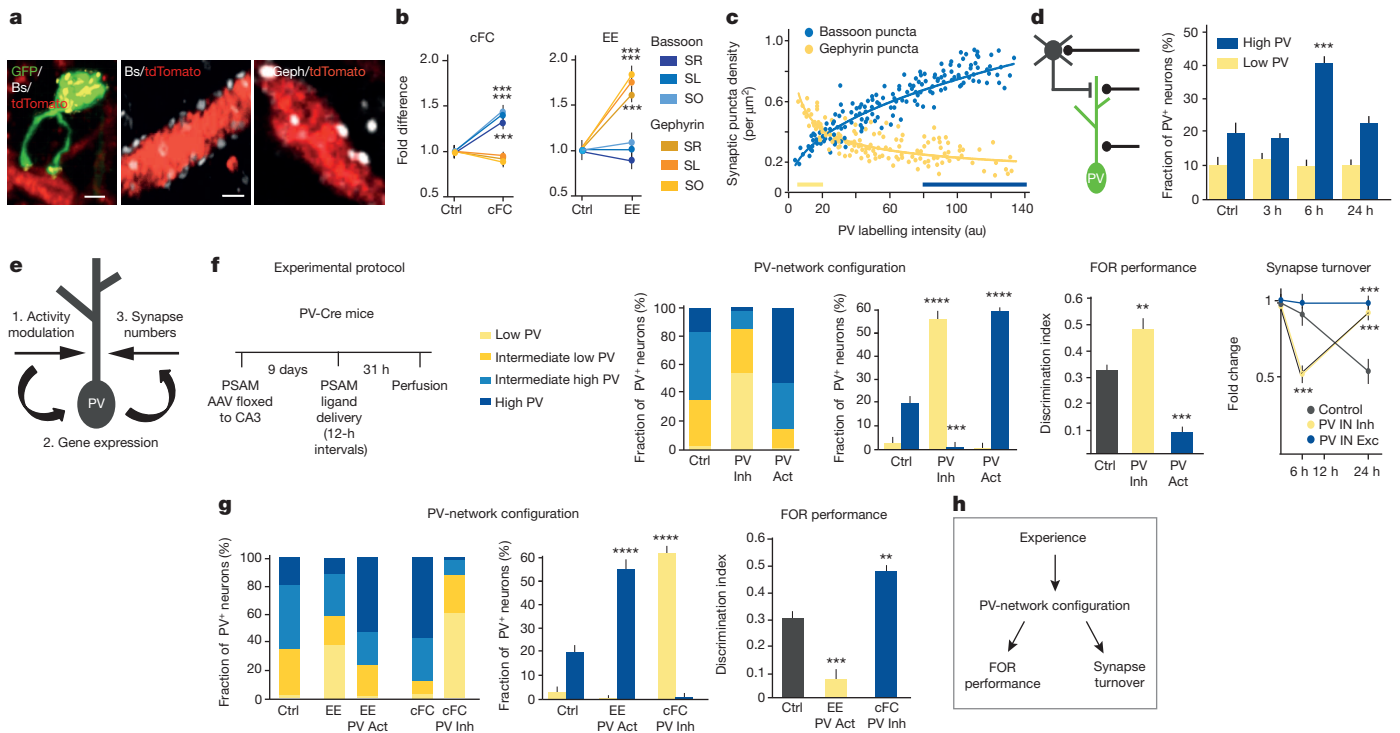


Figure 2 | Configuration shifts in parvalbumin-expressing networks modulate hippocampal synaptic structural plasticity and learning in the adult. **a, b**, Excitatory (Bassoon) and inhibitory (gephyrin) synaptic puncta densities along PV dendrites in CA3b stratum lucidum (SL), radiatum (SR) and oriens (SO). Average values from 5 mice (30 PV-dendrite stretches per mouse) each. Presynaptic inputs onto PB neurons are shown in **a, c**. Local densities of excitatory and inhibitory synaptic puncta as a function of PV immunoreactivity signal. High-PV and low-PV ranges are indicated by the blue and yellow bars. **d**, Optogenetic activation of PV neurons (effect of PV-ChR2 *in vitro* on PV-network configuration); 3 slice cultures and 15–20 PV neurons each. Left, schematic of excitatory (circles) and inhibitory (line) inputs onto PV

synapse turnover at mossy-fibre terminals (Fig. 3c), and first enhanced and then reduced FOR performance (Fig. 3d). In control experiments, mice subjected to the same protocol but in the absence of a learning goal ('swim control'; no hidden platform) exhibited no shifts in hippocampal PV-network configuration (Fig. 3b).

The first day of MWM training, which consisted of four 60-s sessions with visible platform, was sufficient to induce a shift to low-PV-network configuration and improved FOR retrieval on day 2 (Figs 3b, d–f). The low-PV-network shift reversed to baseline after seven but not after three training-free days, in close correspondence with FOR performance and the effectiveness of maze learning²² (Figs 3e, f). To determine whether the shift to high-PV-network configuration at the end of MWM learning reflects the establishment of a spatial reference memory, or rather an absence of spatial prediction errors as mice have learned the spatial map, we carried out MWM experiments in which the relative positions of orientation cues were altered daily from day 5 of training (Fig. 3g). In spite of the match–mismatch regime, mice learned to rapidly locate the hidden platform (Fig. 3h), and acquired a spatial reference memory (Fig. 3i). Notably, however, the PV-neuron network in CA3 remained in the low-PV configuration characteristic of the maze learning phase (Fig. 3j), and the mice continued to exhibit a high performance in the FOR task (Fig. 3k). Taken together, these results suggested that hippocampal CA3 induces and maintains a low-PV-network configuration as long as a clear mismatch is detected between task-related prediction and actual experience, and that it switches to a high-PV-network configuration once predicted and real representations are consistently matched.

To investigate circuit mechanisms that regulate shifts in PV-network configuration during maze learning, we analysed synaptic puncta distributions onto CA3 PV neurons as mice learned the MWM. The shift

neuron (in green). **e**, Schematic of how synaptic activity might induce alterations in gene expression and synapse densities onto PV neurons.

f, Shifting hippocampal PV-interneuron (IN) network configurations through pharmacogenetic activation (Act) and inhibition (Inh) modulates synaptic structural plasticity and learning. $n = 6$ mice each (50 neurons and 40 mossy fibre terminals per mouse). **g, h**, Shifting hippocampal PV network configurations through pharmacogenetic activation and inhibition reverses the effects of EE and cFC on synaptic structural plasticity and learning. $n = 6$ mice each (50 neurons and 40 mossy-fibre terminals per mouse). Colours in **g** are the same as for **f**. Values are means \pm s.e.m. Scale bars, 2 μ m (**a**, left panel); 1 μ m (**a**, middle and right panels). ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

to a low-PV-network was accompanied by an increase in inhibitory synaptic puncta densities onto PV neurons, and the shift to a high-PV-network configuration was accompanied by an increase in excitatory puncta densities (Fig. 4a). However, in marked contrast to environmental enrichment and contextual fear conditioning, increases in synaptic puncta density after MWM learning were highly input-specific. Thus, VIP-positive bouton densities in stratum radiatum specifically increased together with the low-PV-network shift, and stratum lucidum excitatory puncta densities (mossy-fibre input) specifically increased with the high-PV-network shift (Fig. 4a and Supplementary Discussion).

We next investigated the role of VIP neurons in regulating shifts in PV-network configuration and learning during maze navigation. Direct optogenetic stimulation of VIP neurons *in vitro* was sufficient to induce a transient low-PV-network configuration in hippocampal slice cultures (Fig. 4b), providing evidence that VIP neurons can regulate PV-neuron configurations. *In vivo*, transient inhibition of VIP neurons or activation of PV neurons through targeted channel expression and ligand delivery at 12-h intervals, 6 h outside of MWM learning time windows induced a robust shift to a high-PV-network configuration, prevented the shift to a low-PV-network configuration associated with maze learning and suppressed maze learning (Fig. 4c). The pharmacological channel ligand has been shown to be active during the first 1.5–2 h after delivery²⁴, suggesting that its effects on MWM learning were due to shifts in PV network configuration and not due to acute effects on PV- or VIP-neuron activity during behaviour. In complementary conventional pharmacology experiments, VIP greatly increased low-PV-neuron fractions during MWM learning, whereas a VIP receptor antagonist prevented the shift to a low-PV-network configuration (Fig. 4d). Similarly, VIP-treated mice learned more rapidly to locate the

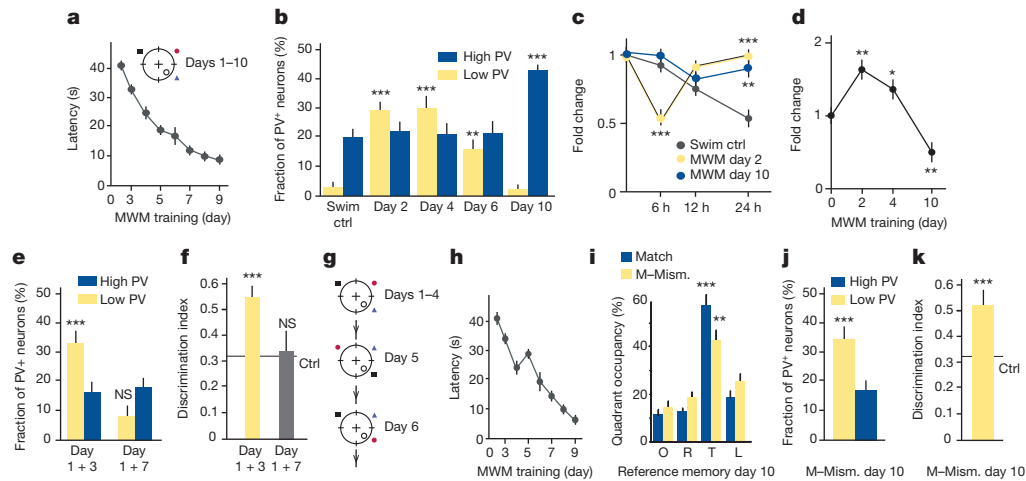


Figure 3 | Maze trial-and-error learning involves early PV-network shift to low-PV configuration, and shift to high-PV configuration upon learning completion. **a–d**, MWM learning curve (**a**; symbols indicate spatial cues), PV-network configuration in CA3b during MWM learning (**b**), and corresponding synaptic structural plasticity (**c**) and FOR performances (**d**). $n = 6–10$ (**a**, **b**, **d**) and 4 (**c**) mice each. **e**, **f**, PV-network shift and enhanced learning upon 1 day of MWM training. Horizontal bar, value in cage control mice. $n = 4$ mice per time point and condition. **g–k**, Persistence of low-PV hidden platform, and already established a reference memory on day 4 (Fig. 4d and Extended Data Fig. 9a). By contrast, mice treated locally with VIP antagonist were severely impaired in learning to navigate the MWM (Fig. 4d), and failed to establish a reference memory after 10 days of training (Extended Data Fig. 9a).

configuration in spite of spatial learning when external cues are rearranged daily, from training day 5 onwards. Schematic of match–mismatch protocol indicating that the spatial cues were shifted daily, which started from day 5 (**g**), corresponding learning curve (**h**), reference spatial memories on day 10 (**i**), PV-network configuration on day 10 for match–mismatch (M–Mism.) protocol (**j**), and corresponding FOR performance (**k**). Values are means \pm s.e.m. $n = 8$ mice. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

To determine whether learning-related shifts in PV-network configuration might be generally associated with trial-and-error learning, we investigated PV neurons in primary motor cortex M1 as mice learned to stay on an accelerating rotating rod. Like the MWM, this motor-skill task involves incremental learning during several days (Fig. 4e), but

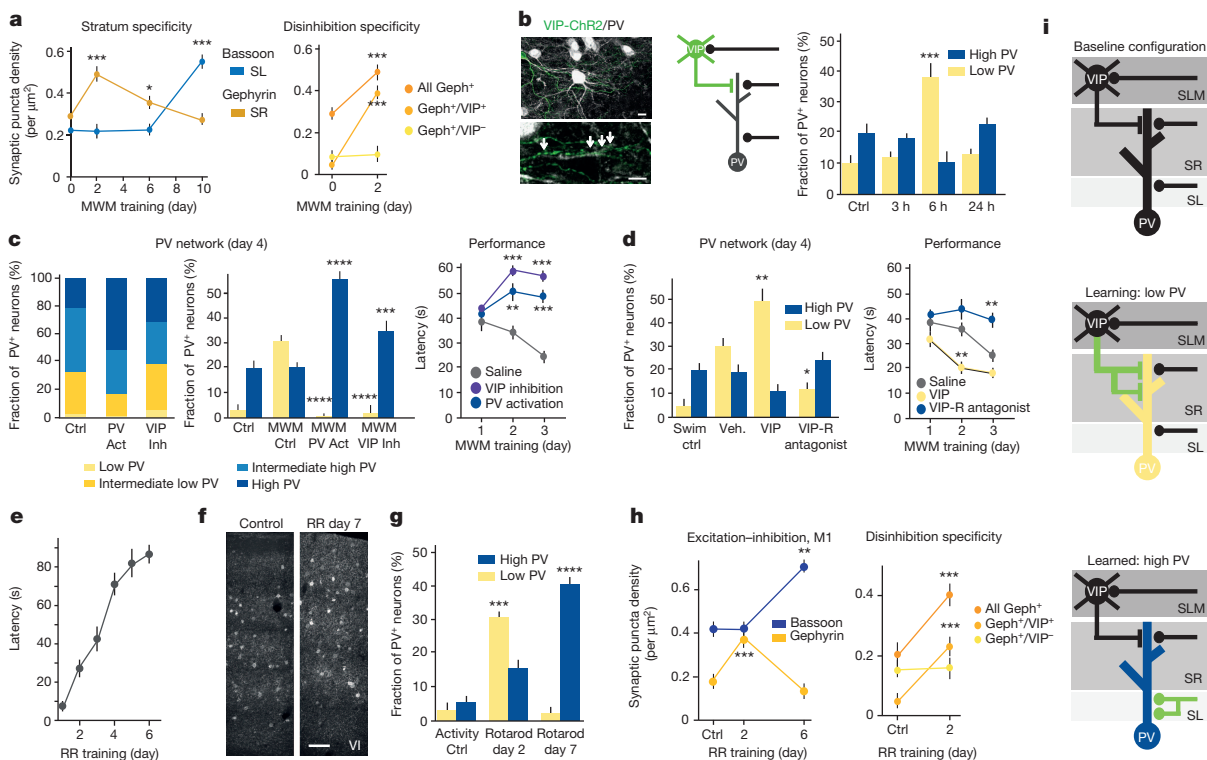


Figure 4 | A PV-VIP microcircuit controls PV-network configuration and behavioural performance during incremental learning. **a**, Specific connectivity rearrangements onto PV neurons during MWM learning. $n = 4$ mice and 15–20 PV neurons each. **b**, Optogenetic activation of VIP-positive interneurons in slice cultures (effect of PV-ChR2). $n = 3$ slice cultures and 15–20 PV neurons each. **c**, Pharmacogenetic activation (PV neurons) or inhibition (VIP neurons) during MWM learning. $n = 6$ mice each (50 neurons per mouse). **d**, VIP-interneuron modulation of hippocampal PV network and behavioural performance. $n = 4$ mice and 20 neurons per mouse each.

Veh., vehicle; VIP-R, VIP receptor. **e–g**, Learning curve (rotarod performance) (**e**), PV immunoreactivity in primary motor cortex (M1) (**f**), and PV-network configuration in M1 layer 2/3 (**g**) during rotarod learning. $n = 6$ (activity Ctrl), 6 (2 days), 8 (7 days). **h**, Specific connectivity rearrangements onto PV neurons during rotarod learning. $n = 4$ mice and 15–20 PV neurons each. **i**, Model of how incremental learning involves early VIP-neuron-induced low-PV network configuration, followed by excitation-mediated high-PV network configuration. Values are means \pm s.e.m. Scale bars, 10 μ m (**b**) and 80 μ m (**f**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

unlike the MWM it involves cerebellum and motor cortex^{17,18}, and not the hippocampus. Rotarod learning produced a major shift to a low-PV-network configuration in M1 during the learning phase, and a large shift to a high-PV-network as the learning curve became saturated after 7 days of training (Fig. 4f, g). As in the hippocampus, the shifts in PV-network were associated with alterations first in inhibitory synaptic puncta densities onto PV dendrites (low-PV configuration) and then with an increase in excitatory synaptic puncta densities (high-PV configuration; Fig. 4h). Notably, and as in the hippocampus, the increase in inhibitory synaptic puncta during learning was specific for VIP-positive boutons (Fig. 4h). The learning-related shifts in PV network and plasticity were highly system-specific: MWM (hippocampal) learning did not induce PV neuron changes in M1, whereas motor learning induced no alterations in PV neurons or mossy-fibre structural plasticity in the hippocampus, and no changes in hippocampus-dependent FOR (Extended Data Fig. 10a–d).

Our results reveal a specific circuit mechanism through which recent experience modulates plasticity, memory consolidation and retrieval, and learning in the adult (Extended Data Fig. 1). We provide evidence that PV basket cells, whose differentiation regulates critical period plasticity^{1–7}, exhibit pronounced differentiation plasticity in the adult, and that this involves major structural alterations in excitatory and inhibitory synaptic puncta densities²⁵ onto PV neurons (Fig. 4i). These alterations might facilitate excitatory recruitment of high-PV- and high-GAD67-expressing basket cells and reduce excitatory recruitment of low-PV and low-GAD67-expressing basket cells. The shift to a low-PV-network configuration might facilitate the combination, consolidation and retrieval of associations established by comparatively weak synaptic networks during incremental trial-and-error learning, whereas the shift to a high-PV-network configuration might promote the establishment of strong memories by comparatively strong synaptic networks and prevent establishment and retrieval of distractor memories. In addition, shifts to low-PV-network configurations might prevent premature establishment of dominant memories, whereas high-PV-network shifts might strengthen appropriate memories and enhance feedforward entrainment upon skill learning due to enhanced gamma coherence^{7,26–28}. Reductions in PV and GAD67 levels have been consistently detected in patients with schizophrenia²⁹, suggesting that dysfunction of excitation–inhibition microcircuits involving PV-network regulation may underlie cognitive disruption in this devastating psychiatric condition.

METHODS SUMMARY

Behavioural experiments were carried out with male mice (postnatal day 55 (P55)–P75 at onset), as described previously^{12,20,22}.

For the PV- and GAD67-intensity analysis, a comparative analysis revealed a cluster of low PV intensities (threshold: 800 arbitrary confocal units (au)) that were present early during development (P15), absent in adult control samples (P60), and induced after environmental enrichment. The 800-au threshold was adopted to classify PV neurons into four subclasses as follows: low PV, 0–800 au; intermediate–low PV, 800–1,600 au; intermediate–high PV, 1,600–2,400 au; high PV, >2,400 au.

For pharmacology, local treatments were carried out on consecutive days, through implanted cannulas (33G, Bilaney Consultants).

For direct local PV- or VIP-neuron control *in vivo* we delivered floxed PSAM-carrying AAV9 bilaterally in dorsal, intermediate and ventral hippocampus in 55- to 70-day-old PV-Cre or VIP-Cre male mice. The agonist PSEM³⁰⁸ was injected intraperitoneally at 12-h intervals at the concentration of 5 $\mu\text{g g}^{-1}$ of animal weight. Behavioural training or testing was performed 6 h post injection.

Unless otherwise stated, statistical groups were compared using unpaired, non-parametric Student's *t*-test. One-way analysis of variance (ANOVA) was used for comparisons of multiple samples. Average values are expressed as means \pm s.e.m.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 May; accepted 5 November 2013.

- Huang, Z. J. *et al.* BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell* **98**, 739–755 (1999).

- Fagioli, M. *et al.* Specific GABA_A circuits for visual cortical plasticity. *Science* **303**, 1681–1683 (2004).
- Sugiyama, S. *et al.* Experience-dependent transfer of Otx2 homeoprotein into the visual cortex activates postnatal plasticity. *Cell* **134**, 508–520 (2008).
- Maya Vetencourt, J. F. *et al.* The antidepressant fluoxetine restores plasticity in the adult visual cortex. *Science* **320**, 385–388 (2008).
- Yazaki-Sugiyama, Y., Kang, S., Câteau, H., Fukai, T. & Hensch, T. K. Bidirectional plasticity in fast-spiking GABA circuits by visual experience. *Nature* **462**, 218–221 (2009).
- Levelt, C. N. & Hübener, M. Critical-period plasticity in the visual cortex. *Annu. Rev. Neurosci.* **35**, 309–330 (2012).
- Sale, A., Berardi, N., Spolidoro, M., Baroncelli, L. & Maffei, L. GABAergic inhibition in visual cortical plasticity. *Front. Cell. Neurosci.* **4**, 10 (2010).
- Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459**, 663–667 (2009).
- Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
- Hendry, S. H. & Jones, E. G. Activity-dependent regulation of GABA expression in the visual cortex of adult monkeys. *Neuron* **1**, 701–712 (1988).
- Welker, E., Soriano, E., Dörfl, J. & Van der Loos, H. Plasticity in the barrel cortex of the adult mouse: transient increase of GAD-immunoreactivity following sensory stimulation. *Exp. Brain Res.* **78**, 659–664 (1989).
- Bednarek, E. & Caroni, P. β -Adducin is required for stable assembly of new synapses and improved memory upon environmental enrichment. *Neuron* **69**, 1132–1146 (2011).
- Caroni, P., Donato, F. & Müller, D. Structural plasticity upon learning: regulation and functions. *Nature Rev. Neurosci.* **13**, 478–490 (2012).
- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Acasády, L., Arabadzisz, D. & Freund, T. F. Correlated morphological and neurochemical features identify different subsets of vasoactive intestinal polypeptide-immunoreactive interneurons in rat hippocampus. *Neuroscience* **73**, 299–315 (1996).
- Dávid, C., Schleicher, A., Zuschratter, W. & Staiger, J. F. The innervation of parvalbumin-containing interneurons by VIP-immunopositive interneurons in the primary somatosensory cortex of the adult rat. *Eur. J. Neurosci.* **25**, 2329–2340 (2007).
- Xu, T. *et al.* Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462**, 915–919 (2009).
- Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
- Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I. & Huang, J. Z. & Kepecs, A. Cortical interneurons that specialize in disinhibitory control. *Nature* **503**, 521–524 (2013).
- Ruediger, S. *et al.* Learning-related feedforward inhibitory connectivity growth required for memory precision. *Nature* **473**, 514–518 (2011).
- Chattopadhyaya, B. *et al.* GAD67-mediated GABA synthesis and signaling regulate inhibitory synaptic innervation in the visual cortex. *Neuron* **54**, 889–903 (2007).
- Ruediger, S., Spirig, D., Donato, F. & Caroni, P. Goal-oriented searching mediated by ventral hippocampus early in trial-and-error learning. *Nature Neurosci.* **15**, 1563–1571 (2012).
- Kleinlogel, S. *et al.* Ultra light-sensitive and fast neuronal activation with the Ca²⁺-permeable channelrhodopsin CatCh. *Nature Neurosci.* **14**, 513–518 (2011).
- Magnus, C. J. *et al.* Chemical and genetic engineering of selective ion channel–ligand interactions. *Science* **333**, 1292–1296 (2011).
- Keck, T. *et al.* Loss of sensory input causes rapid structural changes of inhibitory neurons in adult mouse visual cortex. *Neuron* **71**, 869–882 (2011).
- Volman, V., Behrens, M. M. & Sejnowski, T. J. Downregulation of parvalbumin at cortical GABA synapses reduces network gamma oscillatory activity. *J. Neurosci.* **31**, 18137–18148 (2011).
- Lee, S. H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
- Komiyama, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**, 1182–1186 (2010).
- Lewis, D. A., Curley, A. A., Glausier, J. R. & Volk, D. W. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci.* **35**, 57–67 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Arber for valuable comments on the manuscript and S. Sternson (Janelia), S. Arber and A. Dayer for reagents. F.D. was supported by the NCCR Synapsy. The Friedrich Miescher Institut is part of and supported by the Novartis Research Foundation.

Author Contributions F.D. devised, carried out and analysed all experiments; S.B.R. provided reagents and assistance for the optogenetics experiments; P.C. helped to devise the experiments and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (caroni@fmi.ch).

Distinct fibroblast lineages determine dermal architecture in skin development and repair

Ryan R. Driskell^{1,2}, Beate M. Lichtenberger^{1,2*}, Esther Hoste^{2,3*}, Kai Kretschmar^{1,2*}, Ben D. Simons⁴, Marika Charalambous⁵, Sacri R. Ferron⁵, Yann Herault⁶, Guillaume Pavlovic⁶, Anne C. Ferguson-Smith⁵ & Fiona M. Watt²

Fibroblasts are the major mesenchymal cell type in connective tissue and deposit the collagen and elastic fibres of the extracellular matrix (ECM)¹. Even within a single tissue, fibroblasts exhibit considerable functional diversity, but it is not known whether this reflects the existence of a differentiation hierarchy or is a response to different environmental factors. Here we show, using transplantation assays and lineage tracing in mice, that the fibroblasts of skin connective tissue arise from two distinct lineages. One forms the upper dermis, including the dermal papilla that regulates hair growth and the arrector pili muscle, which controls piloerection. The other forms the lower dermis, including the reticular fibroblasts that synthesize the bulk of the fibrillar ECM, and the preadipocytes and adipocytes of the hypodermis. The upper lineage is required for hair follicle formation. In wounded adult skin, the initial wave of dermal repair is mediated by the lower lineage and upper dermal fibroblasts are recruited only during re-epithelialization. Epidermal β -catenin activation stimulates the expansion of the upper dermal lineage, rendering wounds permissive for hair follicle formation. Our findings explain why wounding is linked to formation of ECM-rich scar tissue that lacks hair follicles^{2–4}. They also form a platform for discovering fibroblast lineages in other tissues and for examining fibroblast changes in ageing and disease.

At embryonic day (E)12.5 mouse epidermis comprises one or two cell layers and the dermis appears homogeneous in composition (Fig. 1a)⁵. By E18.5 the epidermis is fully stratified, contains hair follicles, each with an associated dermal papilla, and is in the process of forming an upper (papillary) dermis that is distinguishable from the lower (reticular) dermis because of its higher cellular density (Fig. 1a). By postnatal day (P)2 the hypodermis has formed, comprising differentiated adipocytes and preadipocytes (Fig. 1a). We identified markers of different fibroblast subpopulations at each developmental stage, on the basis of previous studies^{6,7} and the availability of antibodies for live cell sorting.

The pan-fibroblast marker platelet-derived growth factor receptor- α (Pdgfr- α) is expressed in both the upper and lower dermis at all stages of development (Fig. 1b)⁸. By contrast, there were temporal and spatial changes in the expression of 18 other fibroblast markers (Fig. 1b, c and Extended Data Figs 1–3). From E16.5, CD26 (also known as Dpp4) and B-lymphocyte-induced maturation protein 1 (Blimp1, also known as Prdm1) were selectively expressed in the upper dermis, whereas Sca1 (also known as Ly6a) was selectively expressed in the lower dermis. Delta-like homologue 1 (Dlk1) and leucine-rich repeats and immunoglobulin-like domains protein 1 (Lrig1) were expressed throughout the dermis at E12.5, with Dlk1 expression persisting in the lower dermis

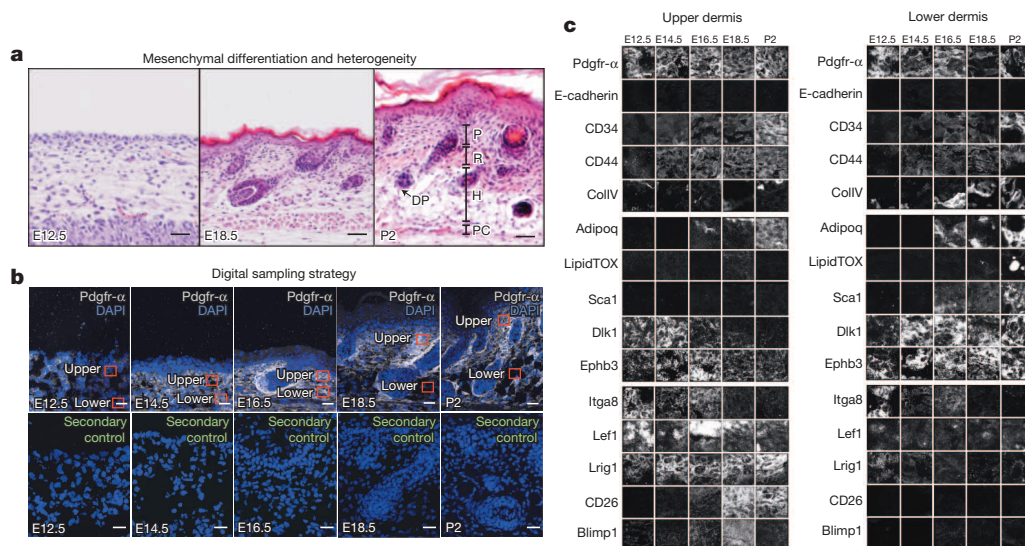


Figure 1 | Morphological and molecular markers of embryonic and postnatal fibroblasts. **a–c**, Sections of mouse back skin. **a**, Haematoxylin and eosin staining. DP, dermal papilla; H, hypodermis; P, papillary dermis; PC, panniculus carnosus; R, reticular dermis. **b**, 10- μ m sections immunostained with Pdgfr- α antibodies and secondary antibody control with DAPI

(4',6-diamidino-2-phenylindole) nuclear counterstain. Red squares show areas digitally sampled for the tissue screen. **c**, Tissue screen of upper and lower dermis. Three biological replicates for each stain were performed. Images are representative samples of sections shown in Extended Data Figs 1 and 2. ColIV, type IV collagen. Scale bars, 50 μ m.

¹Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Cambridge CB2 1QR, UK. ²Centre for Stem Cells and Regenerative Medicine, King's College London, 28th floor, Tower Wing, Guy's Hospital, London SE1 9RT, UK. ³Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ⁴Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. ⁵Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK. ⁶Institut Clinique de la Souris, Parc d'Innovation, 67404 Illkirch-Graffenstaden, Cedex, France.

*These authors contributed equally to this work.

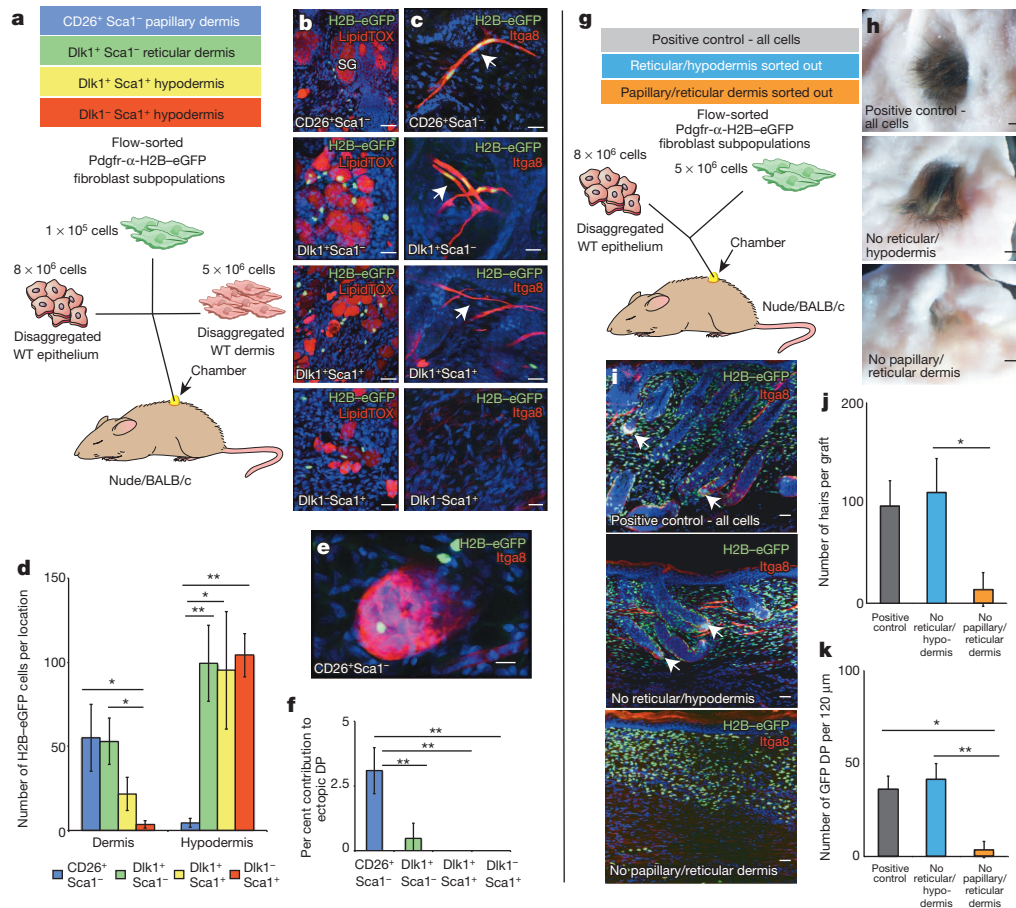


Figure 2 | Skin-reconstitution assays. **a**, Experimental set-up for **b–f**. **b, c**, Grafts immunostained with DAPI counterstain (blue). SG, sebaceous gland. Arrows, APM. **d–f**, Contribution of Pdgfr-α-H2B-eGFP cells to dermal compartments. DP, dermal papilla. **g**, Experimental set-up for **h–k**. **h**, Macroscopic views of grafts. **i**, Contribution of GFP⁺ cells to grafts. Arrows,

GFP⁺ dermal papilla. **j**, Hair follicles per graft. **k**, GFP⁺ dermal papillae per 120-μm graft. *n* = 3 biological replicates per experiment. **P* < 0.05, ***P* < 0.005. Horizontal whole mounts shown. Scale bars, 40 μm (**b, c**), 30 μm (**e**), 2.5 mm (**h**), 50 μm (**i**).

from E18.5 and *Lrig1* expression persisting in the upper dermis. The changes in abundance of CD26⁺, Sca1⁺ and Dlk1⁺ cells were confirmed by flow cytometry of green fluorescent protein (GFP)⁺ dermal cells from Pdgfr-α-H2B-eGFP mice⁸ (in which enhanced GFP (eGFP) is fused to histone H2B) (Extended Data Fig. 3). Several of the markers were also differentially expressed in neonatal human skin (Extended Data Fig. 4).

To evaluate the differentiation potential of different dermal fibroblasts (Fig. 2a), cells were flow sorted from P2 Pdgfr-α-H2B-eGFP dermis⁸ (Extended Data Fig. 5c–e), combined with unlabelled epidermal and dermal cells and injected into chambers implanted into nude BALB/c mice⁹. Papillary dermal cells (CD26⁺ Sca1⁻) contributed exclusively to the upper dermis (Fig. 2b–d and Extended Data Fig. 6a), including the dermal papilla and arrector pili muscle (APM) (Fig. 2e, f). Hypodermal fibroblasts (Dlk1⁺ Sca1⁺ and Dlk1⁻ Sca1⁺) differentiated into adipocytes but not into APM or dermal papilla (Fig. 2b–d and Extended Data Fig. 6c, d). Dlk1⁺ Sca1⁻ cells, located primarily at the boundary between the reticular dermis and hypodermis (Extended Data Fig. 5a, b), contributed to all dermal mesenchymal compartments (Fig. 2b–d and Extended Data Fig. 6b). However, although we cannot rule out the existence of multipotent fibroblasts in adult skin, Dlk1⁺ Sca1⁻ cells did not persist after P10, when Dlk1 was no longer expressed in the dermis (Extended Data Fig. 4d and data not shown).

The number of hair follicles was higher in grafts containing upper dermal (CD26⁺ Sca1⁻) cells than Dlk1⁻ Sca1⁺ hypodermal cells (Extended Data Fig. 6e–i). Dlk1⁻ Sca1⁺ cells are a subpopulation of Lin⁻ CD34⁺ CD29⁺ Sca1⁺ adipocyte precursors, which also express Pdgfr-α (Extended Data Fig. 5f–i)¹⁰. To establish whether hypodermal

cells were less effective at supporting hair follicle formation we compared control grafts containing unfractionated P2 Pdgfr-α-H2B-eGFP⁺ cells with grafts in which we excluded papillary/reticular cells (CD26⁺, Dlk1⁺ and Sca1⁻ cells depleted) or reticular/hypodermal cells (Dlk1⁺ Sca1⁺ cells depleted) (Fig. 2g–k and Extended Data Fig. 6j–l). As CD26, Dlk1 and Sca1 are not expressed in the dermal papilla at this stage (data not shown) all grafts contained dermal papilla cells. The contribution of GFP⁺ cells to the grafts was extensive in all cases (Fig. 2i). Hair follicle formation was similar in the grafts of unfractionated dermal cells and those depleted of reticular and hypodermal cells. By contrast, when the papillary and reticular dermal cells were excluded very few hair follicles formed (Fig. 2h–k). Thus P2 skin contains cells that are restricted to forming either the upper or lower dermal lineages on skin reconstitution, the upper dermal cells being required for hair follicle formation.

We next performed lineage tracing with specific promoters. We crossed Dlk1-CreER mice with Rosa-CAG-loxP-stop-loxP (LSL)-tdTomato mice (hereafter LSL-tdTomato mice, in which tdTomato is expressed when the stop codon is removed via Cre-mediated recombination), then labelled with tamoxifen at E12.5 (when all fibroblasts are Dlk1⁺; Fig. 1c) or E16.5 (when only reticular fibroblasts express Dlk1; Fig. 1c) and collected at P2 or P21 (Fig. 3a–k). Flow cytometry established that the labelling accuracy was over 95% (Extended Data Fig. 7a–c). When Dlk1-expressing cells were labelled at E12.5, all of the dermal mesenchymal compartments (Pdgfr-α⁺) were labelled (Fig. 3a–d). By contrast, when cells were labelled at E16.5 the only labelled fibroblasts were reticular and hypodermal cells, including mature adipocytes (Fig. 3e–k).

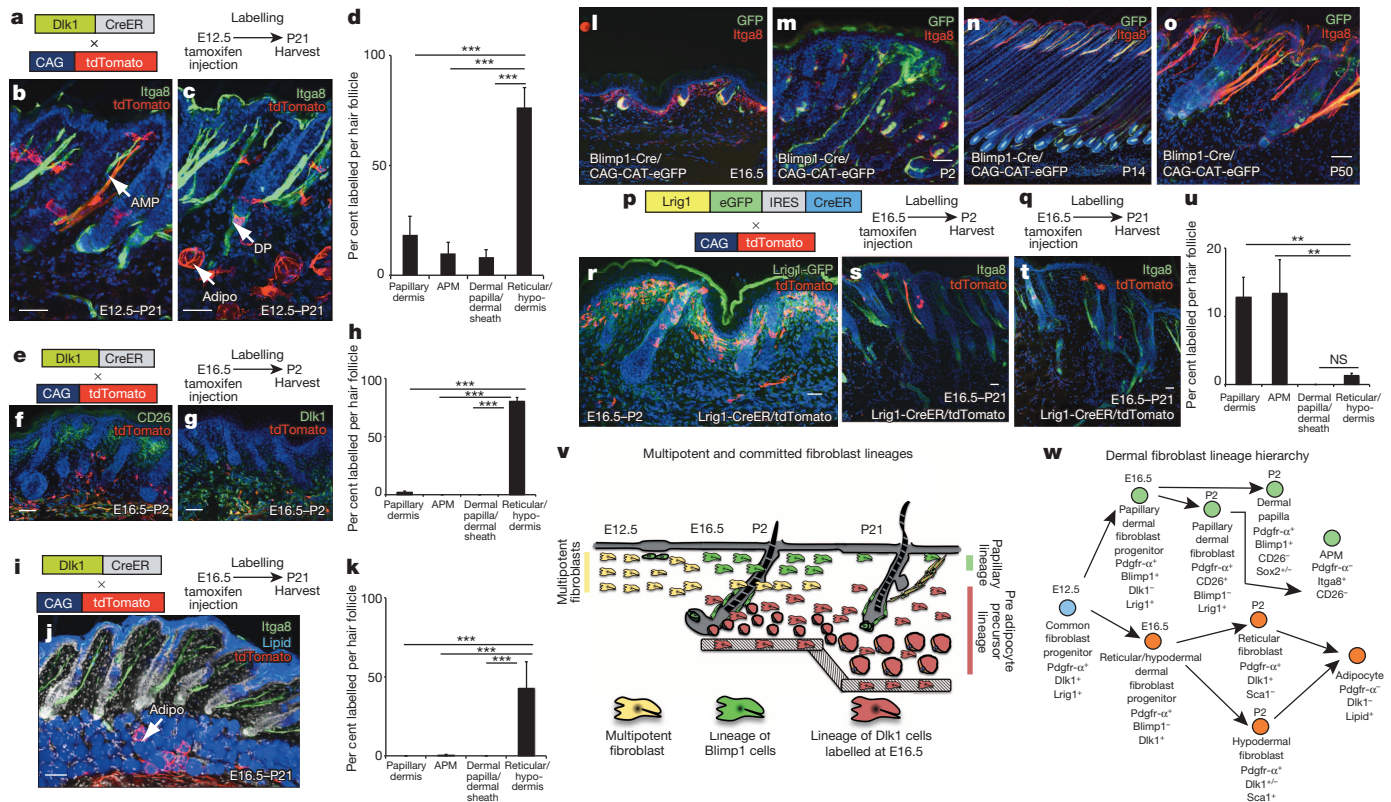


Figure 3 | Fibroblast lineage commitment during skin development.

a, e, i, p, q. Labelling strategies. **b, c, f, g, j, l–o, r–t.** Contribution of labelled cells to different dermal compartments. Adipo, adipocyte; DP, dermal papilla. **d, h, k, u.** Average number of labelled cells per unit area of dermis (defined by

hair follicle spacing). **v, w.** Schematic summary of lineage restriction. $n = 3$ biological replicates analysed per experiment. Sections counterstained with DAPI. NS, not significant. ** $P < 0.005$, *** $P < 0.0005$. Scale bars, 50 μ m.

Therefore after E16.5 dermal fate restriction occurs, such that cells expressing Dlk1 only give rise to the lower dermal lineages. We confirmed this by examining adipogenesis of dermal cells in culture (Extended Data Fig. 3d–g; see Supplementary Information)^{6,11}.

We next performed lineage tracing with Blimp1-Cre. Between E16.5 and E18.5, Blimp1 was expressed in the dermal papilla, dermal sheath and papillary dermis (Fig. 1c and Extended Data Fig. 7d, e)^{12,13}, but from P2 through adulthood dermal Blimp1 expression was confined to the dermal papilla and dermal sheath (Extended Data Fig. 7f, g). Blimp1-expressing epidermal and endothelial cells (Extended Data Fig. 7h, i)^{13–15} were excluded from analysis on the basis of co-expression of relevant markers. The fidelity of the Blimp1 promoter was established by flow sorting (Extended Data Fig. 7j, k).

At all time points, dermal papilla and dermal sheath cells were labelled in Blimp1-Cre \times CAG-CAT-eGFP (hereafter LSL-GFP) mice (Fig. 3l–o). At E16.5 the papillary dermis was unlabelled (Fig. 3l), but at E18.5 the papillary dermis contained GFP⁺ cells, consistent with the pattern of endogenous Blimp1 expression (Extended Data Fig. 7e, h). From P14 onwards all APM were GFP⁺, and only a few GFP⁺ fibroblasts remained in the papillary dermis (Fig. 3n, o). Thus P2 papillary dermal fibroblasts are the progenitors of the APM. Labelled papillary fibroblasts never contributed to the reticular dermis or hypodermis.

We used Lrig1 as an additional marker for lineage tracing¹⁶ (Figs 1c, 3r–u and Extended Data Fig. 7l–q). The specificity of the Lrig1-GFP-IRES-CreER reporter line¹⁷ was confirmed by flow cytometry (Extended Data Fig. 7n). P2 dermal cells from Lrig1-GFP-IRES-CreER \times LSL-tdTomato crosses treated with tamoxifen at E16.5 expressed Pdgfr- α and CD26 but not Sca1 (Extended Data Fig. 7n–q). At P2 labelling was restricted to the papillary dermis (Fig. 3r and Extended Data Fig. 7m). At P21 the APM and papillary dermis were labelled, whereas the reticular dermis and fat were unlabelled (Fig. 3s–u). Our data suggest that loss of papillary dermis after the first hair follicle cycle^{8,18} is due to

differentiation of papillary fibroblasts into the APM, which was further confirmed by clonal analysis in Pdgfr- α -CreER \times LSL-GFP mice (Extended Data Fig. 8). It is interesting that the papillary dermis is more pronounced in human skin, which has fewer hair follicles and therefore fewer APMs, than mouse skin⁵. We conclude that from E16.5 the developing dermis undergoes fate restriction such that cells in the upper dermis that express Lrig1 or Blimp1 give rise to the dermal papilla, APM and papillary fibroblasts, whereas Dlk1-positive cells in the lower dermis give rise to the reticular dermis, hypodermis and adipocyte layer (Fig. 3v, w).

To investigate how fibroblasts of different lineages contribute to skin repair, we created full-thickness wounds in adult back skin. We ruled out any contribution of bone-marrow-derived cells to the Pdgfr- α ⁺ dermal compartment (Extended Data Fig. 9a–k and Supplementary Information). When Pdgfr- α -CreER \times LSL-tdTomato mice that had been injected with tamoxifen at E17.5 were wounded in adulthood, we found tdTomato⁺ cells throughout the repairing dermis (Extended Data Fig. 9l, n). When we examined wound healing in Dlk1-CreER \times LSL-tdTomato mice that had been injected with tamoxifen at E16.5 to selectively label the lower dermis (Fig. 4a–c and Extended Data Fig. 10a, b), a large proportion of wound fibroblasts, including cells that expressed the myofibroblast marker α -smooth muscle actin¹⁹ (Extended Data Fig. 10b), were tdTomato⁺ (Fig. 4b, c). In adult skin of Blimp1-Cre \times LSL-GFP mice only upper dermal fibroblasts and CD31⁺ endothelial cells were GFP⁺ (Fig. 4e). Seven days after wounding no GFP⁺ fibroblasts were present in the wound bed (Fig. 4e and Extended Data Fig. 10c, d) and at day 17 GFP⁺ fibroblasts were exclusively detected immediately underneath the reformed epidermis (Fig. 4f and Extended Data Fig. 10d).

We conclude that the first wave of dermal regeneration depends on cells of the reticular dermis and hypodermis, which are not capable of inducing hair follicles (Fig. 2i–k) and which elaborate the collagenous

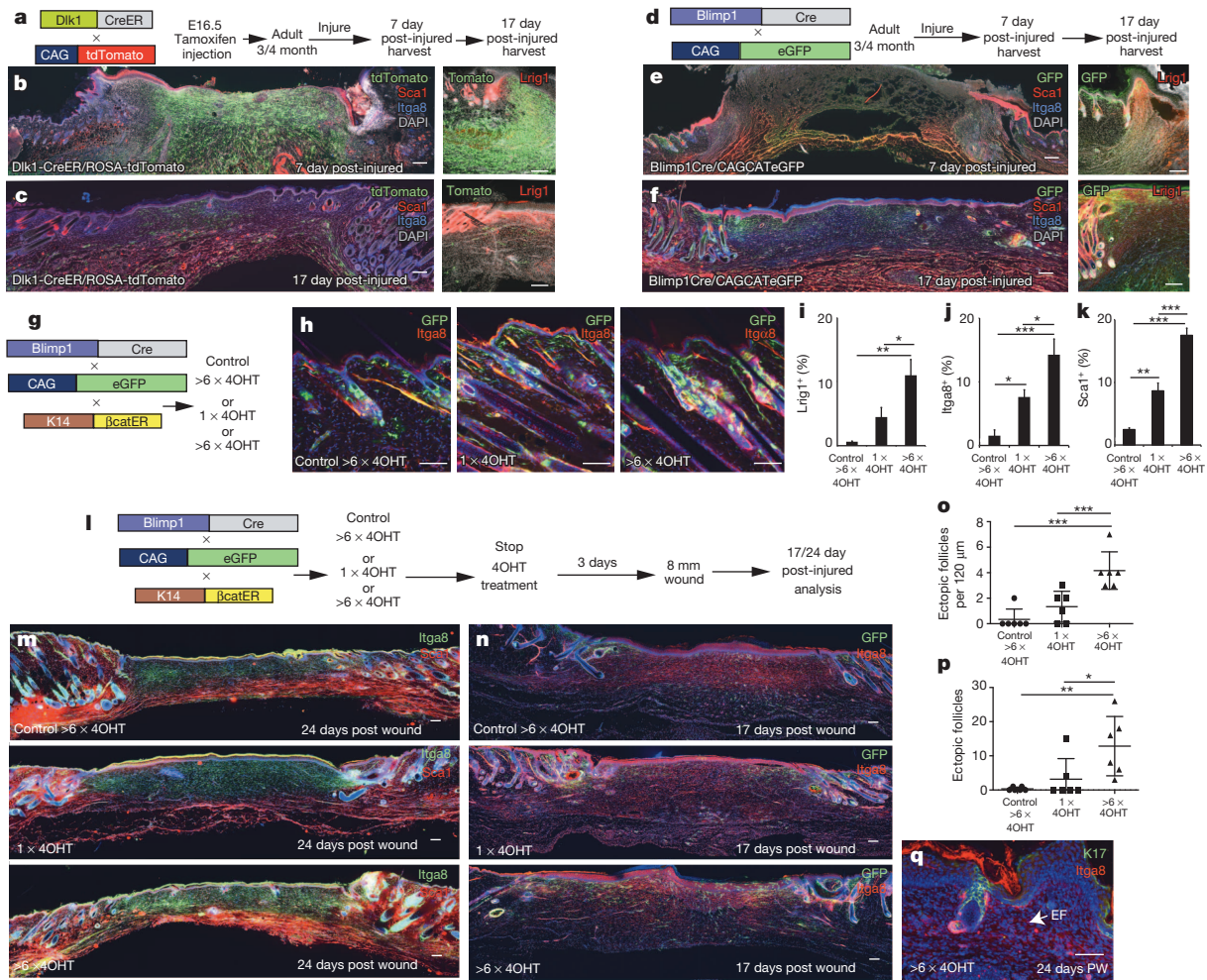


Figure 4 | Contribution of fibroblast lineages in adult skin.
a, d, g, i, Labelling strategies. **b, c, e, f, h, m, n, q**, Histological analysis of horizontal whole mounts. **i–k**, Flow-cytometric quantification of fibroblast markers. **o, p**, Quantification of ectopic hair follicles in 17-day (**o**) and 24-day

wound bed whole-mount sections (**p**). $n = 3$ (**i–k**) and $n = 6$ (**o, p**) biological replicates. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$. Scale bars, 200 μm (**b, c, e, f**), 100 μm (**m, n, q**) or 50 μm (higher magnification views). 4OHT, tamoxifen.

ECM characteristic of fibrosis⁶. The upper dermal lineage is not repopulated until re-epithelialization and contributes exclusively to the papillary dermis. This explains both the rapid reconstitution of the skin adipocyte layer after wounding¹⁹ and the absence of hair follicles in newly closed wounds²⁰.

Sustained activation of the canonical Wnt pathway in adult epidermis induces growth of existing hair follicles (anagen), formation of new hair follicles^{21,22}, fibroblast proliferation and dermal ECM remodelling⁸. To examine the impact on the different dermal lineages we used K14 $\Delta\text{N}\beta$ -cateninER transgenic mice. One application of tamoxifen induces anagen, whereas repeated applications stimulate ectopic hair follicle formation²¹. Epidermal β -catenin activation led to expansion of the upper and lower dermal compartments, the effects being more pronounced in mice that received multiple applications of tamoxifen (Fig. 4h–k and Extended Data Fig. 10e–g). The relative locations of the different dermal cells were unchanged (Extended Data Fig. 9o, p). Expansion of the upper dermal lineage was evident in K14 $\Delta\text{N}\beta$ -cateninER x Blimp1-Cre/LSL-GFP mice (Fig. 4g, h). As epidermal Wnt signalling promotes hair follicle growth, the expansion of the lower dermis could explain why the skin adipocyte layer increases in thickness during anagen²².

To test whether epidermal β -catenin activation before wounding would promote hair follicle formation, K14 $\Delta\text{N}\beta$ -cateninER/Blimp1-Cre/LSL-GFP mice were given repeated applications of tamoxifen and

subsequently wounded (Fig. 4l). This led to increased numbers of Itga8⁺ and Lig1⁺ fibroblasts in the wound bed (Fig. 4m and Extended Data Fig. 10h), an increased number of cells marked by Blimp1-Cre (Fig. 4n) and a pronounced stimulation of hair follicle formation (Fig. 4o–q). We conclude that dermal remodelling in response to epidermal β -catenin activation leads to expansion of the upper and lower dermis, the increase in the upper papillary dermis being permissive for new hair follicle formation. It will be interesting to identify the downstream signalling pathways involved, one candidate being Hedgehog²¹. Wnt signalling regulates cells of the immune system²³ and a role for infiltrating $\gamma\delta$ T cells in promoting hair follicle formation has recently been described²⁴. The Wnt effector transcription factors Lef1 and Tcf3 were differentially expressed by upper and lower dermal fibroblasts (Extended Data Fig. 5d, e), raising the intriguing possibility that different fibroblast lineages differ in their Wnt responsiveness and interactions with inflammatory cells²⁴.

In conclusion, the lineage hierarchies we have described reveal the cellular origins of the heterogeneous architecture of the dermis and provide a mechanistic basis for understanding the changes that take place during repair of adult skin wounds and in response to epidermal Wnt activation. We now have the opportunity to uncover events that result in dermal lineage specification, and discover how the different lineages contribute to the changes that occur during ageing²⁵ and tumour formation¹.

METHODS SUMMARY

The mouse lines, tamoxifen treatments, chamber grafting assay and bone marrow transplantation experiments are described in the Methods^{26–30}. Most experiments were performed on female mice; however, the markers evaluated showed the same distribution in males. 5-µm frozen sections and horizontal whole mounts¹¹ were prepared, stained and analysed as described in the Methods. Flow cytometry of P2 disaggregated dermal preparations was performed as previously described¹¹. Antibodies for immunolabelling are listed in the Methods. RT-PCR was performed as previously described²⁷. Analysis was performed by the delta-Ct method. Primer probe sets were purchased from Applied Biosystems and used according to manufacturer's recommendations.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 February; accepted 15 October 2013.

- Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nature Rev. Cancer* **6**, 392–401 (2006).
- Gurtner, G. C., Werner, S., Barrandon, Y. & Longaker, M. T. Wound repair and regeneration. *Nature* **453**, 314–321 (2008).
- Martin, P. Wound healing—aiming for perfect skin regeneration. *Science* **276**, 75–81 (1997).
- Shah, M., Foreman, D. M. & Ferguson, M. W. Neutralising antibody to TGF- $\beta_{1,2}$ reduces cutaneous scarring in adult rodents. *J. Cell Sci.* **107**, 1137–1157 (1994).
- Van Exan, R. J. & Hardy, M. H. The differentiation of the dermis in the laboratory mouse. *Am. J. Anat.* **169**, 149–164 (1984).
- Dulauroy, S., Di Carlo, S. E., Langa, F., Eberl, G. & Peduto, L. Lineage tracing and genetic ablation of ADAM12⁺ perivascular cells identify a major source of profibrotic cells during acute tissue injury. *Nature Med.* **18**, 1262–1270 (2012).
- Janson, D. G., Saintigny, G., van Adrichem, A., Mahe, C. & El Ghalbzouri, A. Different gene expression patterns in human papillary and reticular fibroblasts. *J. Invest. Dermatol.* **132**, 2565–2572 (2012).
- Collins, C. A., Kretschmar, K. & Watt, F. M. Reprogramming adult dermis to a neonatal state through epidermal activation of β -catenin. *Development* **138**, 5189–5199 (2011).
- Jensen, K. B., Driskell, R. R. & Watt, F. M. Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis. *Nature Protocols* **5**, 898–911 (2010).
- Festa, E. *et al.* Adipocyte lineage cells contribute to the skin stem cell niche to drive hair cycling. *Cell* **146**, 761–771 (2011).
- Driskell, R. R. *et al.* Clonal growth of dermal papilla cells in hydrogels reveals intrinsic differences between Sox2-positive and -negative cells *in vitro* and *in vivo*. *J. Invest. Dermatol.* **132**, 1084–1093 (2012).
- Lesko, M. H., Driskell, R. R., Kretschmar, K., Goldie, S. J. & Watt, F. M. Sox2 modulates the function of two distinct cell lineages in mouse skin. *Dev. Biol.* **382**, 15–26 (2013).
- Robertson, E. J. *et al.* Blimp1 regulates development of the posterior forelimb, caudal pharyngeal arches, heart and sensory vibrissae in mice. *Development* **134**, 4335–4345 (2007).
- Horsley, V. *et al.* Blimp1 defines a progenitor population that governs cellular input to the sebaceous gland. *Cell* **126**, 597–609 (2006).
- Magnúsdóttir, E. *et al.* Epidermal terminal differentiation depends on B lymphocyte-induced maturation protein-1. *Proc. Natl Acad. Sci. USA* **104**, 14988–14993 (2007).
- Gomez, C. *et al.* The interfollicular epidermis of adult mouse tail comprises two distinct cell lineages that are differentially regulated by Wnt, Edaradd, and Lrig1. *Stem Cell Reports* **1**, 19–27 (2013).
- Page, M. E., Lombard, P., Ng, F., Gottgens, B. & Jensen, K. B. The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell* **13**, 471–482 (2013).
- Egawa, G., Osawa, M., Uemura, A., Miyachi, Y. & Nishikawa, S. Transient expression of ephrin b2 in perinatal skin is required for maintenance of keratinocyte homeostasis. *J. Invest. Dermatol.* **129**, 2386–2395 (2009).
- Schmidt, B. A. & Horsley, V. Intradermal adipocytes mediate fibroblast recruitment during skin wound healing. *Development* **140**, 1517–1527 (2013).
- Ito, M. *et al.* Wnt-dependent *de novo* hair follicle regeneration in adult mouse skin after wounding. *Nature* **447**, 316–320 (2007).
- Silva-Vargas, V. *et al.* β -catenin and Hedgehog signal strength can specify number and location of hair follicles in adult epidermis without recruitment of bulge stem cells. *Dev. Cell* **9**, 121–131 (2005).
- Plikus, M. V. *et al.* Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. *Nature* **451**, 340–344 (2008).
- Staal, F. J., Luis, T. C. & Tiemessen, M. M. WNT signalling in the immune system: WNT is spreading its wings. *Nature Rev. Immunol.* **8**, 581–593 (2008).
- Gay, D. *et al.* Fgf9 from dermal $\gamma\delta$ T cells induces hair follicle neogenesis after wounding. *Nature Med.* **19**, 916–923 (2013).
- Giangreco, A., Qin, M., Pintar, J. E. & Watt, F. M. Epidermal stem cells are retained *in vivo* throughout skin aging. *Aging Cell* **7**, 250–259 (2008).
- Arwert, E. N. *et al.* Tumor formation initiated by nondividing epidermal cells via an inflammatory infiltrate. *Proc. Natl Acad. Sci. USA* **107**, 19903–19908 (2010).
- Driskell, R. R., Giangreco, A., Jensen, K. B., Mulder, K. W. & Watt, F. M. Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis. *Development* **136**, 2815–2823 (2009).
- Rivers, L. E. *et al.* PDGFRA/NG2 glia generate myelinating oligodendrocytes and piriform projection neurons in adult mice. *Nature Neurosci.* **11**, 1392–1401 (2008).
- Kawamoto, S. *et al.* A novel reporter mouse strain that expresses enhanced green fluorescent protein upon Cre-mediated recombination. *FEBS Lett.* **470**, 263–268 (2000).
- Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. & Soriano, P. Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. *Mol. Cell Biol.* **23**, 4013–4025 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by the Wellcome Trust (F.M.W., A.C.F.-S.), the Medical Research Council (MRC) (F.M.W., A.C.F.-S.) and the European Union FP7 programme: TUMIC (F.M.W.), HEALING (F.M.W.) and EpigeneSys (A.C.F.-S.). B.M.L. is the recipient of a FEBS long-term fellowship. K.K. is the recipient of a MRC PhD Studentship. The authors acknowledge financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London (KCL) and King's College Hospital NHS Foundation Trust. Input from M. Mastrogiannaki, A. Reimer and B. Trappmann is gratefully acknowledged. The authors also thank J. Connelly and G. Theocharis for providing tissue, KCL core staff, in particular J. Castle, C. Pearce and D. Cooper, for technical support, and J. Harris and the Nikon Imaging Centre at KCL for expert assistance.

Author Contributions R.R.D. and F.M.W. designed the experiments, performed data analysis, interpreted the results and wrote the manuscript. B.M.L., E.H., K.K., A.C.F.-S., S.R.F., B.D.S. and M.C. assisted in performing and designing experiments, analysing data and interpreting results. Y.H. and G.P. generated the Dlk1-CreER transgenic mouse.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.M.W. (fiona.watt@kcl.ac.uk).

Derivation of novel human ground state naive pluripotent stem cells

Ohad Gafni^{1*}, Leehee Weinberger^{1*}, Abed AlFatah Mansour^{1*}, Yair S. Manor^{1*}, Elad Chomsky^{1,2,3*}, Dalit Ben-Yosef^{4,5}, Yael Kalma⁴, Sergey Viukov¹, Itay Maza¹, Asaf Zviran¹, Yoach Rais¹, Zohar Shipony^{2,3}, Zohar Mukamel^{2,3}, Vladislav Krupalnik¹, Mirie Zerbib¹, Shay Geula¹, Inbal Caspi¹, Dan Schneir¹, Tamar Shwartz⁴, Shlomit Gilad⁶, Daniela Amann-Zalcenstein⁶, Sima Benjamin⁶, Ido Amit⁷, Amos Tanay^{2,3}, Rada Massarwa¹, Noa Novershtern¹ & Jacob H. Hanna¹

Mouse embryonic stem (ES) cells are isolated from the inner cell mass of blastocysts, and can be preserved *in vitro* in a naive inner-cell-mass-like configuration by providing exogenous stimulation with leukaemia inhibitory factor (LIF) and small molecule inhibition of ERK1/ERK2 and GSK3 β signalling (termed 2i/LIF conditions)^{1,2}. Hallmarks of naive pluripotency include driving *Oct4* (also known as *Pou5f1*) transcription by its distal enhancer, retaining a pre-inactivation X chromosome state, and global reduction in DNA methylation and in H3K27me3 repressive chromatin mark deposition on developmental regulatory gene promoters³. Upon withdrawal of 2i/LIF, naive mouse ES cells can drift towards a primed pluripotent state resembling that of the post-implantation epiblast⁴. Although human ES cells share several molecular features with naive mouse ES cells⁵, they also share a variety of epigenetic properties with primed murine epiblast stem cells (EpiSCs)^{6,7}. These include predominant use of the proximal enhancer element to maintain OCT4 expression, pronounced tendency for X chromosome inactivation in most female human ES cells, increase in DNA methylation and prominent deposition of H3K27me3 and bivalent domain acquisition on lineage regulatory genes⁷. The feasibility of establishing human ground state naive pluripotency *in vitro* with equivalent molecular and functional features to those characterized in mouse ES cells remains to be defined¹. Here we establish defined conditions that facilitate the derivation of genetically unmodified human naive pluripotent stem cells from already established primed human ES cells, from somatic cells through induced pluripotent stem (iPS) cell reprogramming or directly from blastocysts. The novel naive pluripotent cells validated herein retain molecular characteristics and functional properties that are highly similar to mouse naive ES cells, and distinct from conventional primed human pluripotent cells. This includes competence in the generation of cross-species chimaeric mouse embryos that underwent organogenesis following microinjection of human naive iPS cells into mouse morulas. Collectively, our findings establish new avenues for regenerative medicine, patient-specific iPS cell disease modelling and the study of early human development *in vitro* and *in vivo*.

Naive 2i/LIF conditions are not sufficient to elicit a self-renewal response in human ES cells/human iPS cells and result in their differentiation⁷. Nevertheless, the proof of concept for the bi/metastability between naive and primed pluripotent states in rodents⁸, has raised the possibility that the human genetic background might be less permissive for allowing *in vitro* preservation of ground state-naive pluripotency in comparison to rodents^{1,7}. Further, it is feasible that different exogenous factor combinations, in addition to 2i/LIF, may allow *in vitro* stabilization of transgene-independent and indefinitely stable human naive pluripotent cells¹. To define such conditions, we used a 'secondary' human

C1.2 iPS cell line containing doxycycline (dox)-inducible lentiviral transgenes encoding OCT4, SOX2, KLF4 and c-MYC (OSKM) reprogramming factors, that was targeted with an OCT4-GFP-2A-PURO knock-in reporter construct (Fig. 1a)⁷. As previously established, exogenous expression of reprogramming factors encoding transgenes by dox supplementation allows maintenance of cells that are morphologically similar to mouse ES cells, while retaining approximately 60% OCT4-GFP⁺ cell fraction in 2i/LIF conditions (Extended Data Fig. 1a, b)⁷. This cell line was used to screen for components that, upon dox withdrawal, could stabilize C1.2 human iPS cells in 2i/LIF indefinitely with homogenous OCT4-GFP⁺ expression (in nearly 100% of the cells) (Fig. 1a).

Whereas C1.2 cells rapidly differentiated in 2i/LIF only conditions, the combined action of 16 factor conditions (16F, divided into pool 1 and pool 2 subgroups) attenuated the differentiation propensity and allowed retaining 32% of OCT4-GFP⁺ cells as measured at day 14 after dox withdrawal (Extended Data Fig. 1b). This indicated that the 16F combination contains factors that cooperatively promote human pluripotency maintenance in 2i/LIF conditions. When pool 2 components were removed (Extended Data Fig. 1b), OCT4-GFP⁺ cell fraction increased relative to 16F combination, indicating that pool 2 contained factors that were negatively influencing GFP⁺ cell maintenance (Extended Data Fig. 1b). We propose that FGFR and TGFR pathway inhibition was detrimental to growing human pluripotent cells in 2i/LIF conditions, as FGF2 and TGF- β cytokines have an evolutionary divergent function in promoting pluripotency in humans by inducing naive pluripotency KLF4 and NANOG transcription factor expression in human ES cells⁹, but not in murine EpiSCs, where they promote murine pluripotency priming. Indeed, removing both TGFR and FGFR inhibitors recapitulated the phenotype obtained when removing dox and pool 2 components (Extended Data Fig. 1b). Moreover, pool 1 components supplemented with exogenous FGF2 and TGF- β 1 cytokines resulted in homogenous OCT4-GFP detection in >95% of cells independent of dox in 2i/LIF containing conditions (Extended Data Fig. 1b). We then tested which of the pool 1 components were essential, and found that 2i/LIF, p38i, JNKi together with FGF2 and TGF- β 1 cytokine supplementation were essential to maintain exogenous transgene-independent GFP⁺ C1.2 clones (Extended Data Fig. 1c). Secondary optimization identified Rho-associated coiled-coil kinases (ROCK)¹⁰ and protein kinase C (PKC)¹¹ inhibitors (included in pool 2, Fig. 1a) as optional boosters of naive cell viability and growth (Extended Data Fig. 1d–f), and resulted in optimized chemically defined conditions termed NHSM (naive human stem cell medium) (Fig. 1b). NHSM conditions enabled expansion of karyotypically normal OCT4-GFP⁺ C1.2 human iPS cells for over 50 passages independent of exogenous transgene activation (Fig. 1c). Remarkably, human pluripotent cells grown

¹The Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. ²The Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. ³The Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. ⁴Wolfe PGD Stem Cell Lab, Racine IVF Unit, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, Tel-Aviv, Israel. ⁵The Department of Cell and Developmental Biology, Sackler Medical School, Tel-Aviv University, Israel. ⁶The Israel National Center for Personalized Medicine (INCPM), Weizmann Institute of Science, Rehovot 76100, Israel. ⁷The Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel.

*These authors contributed equally to this work.

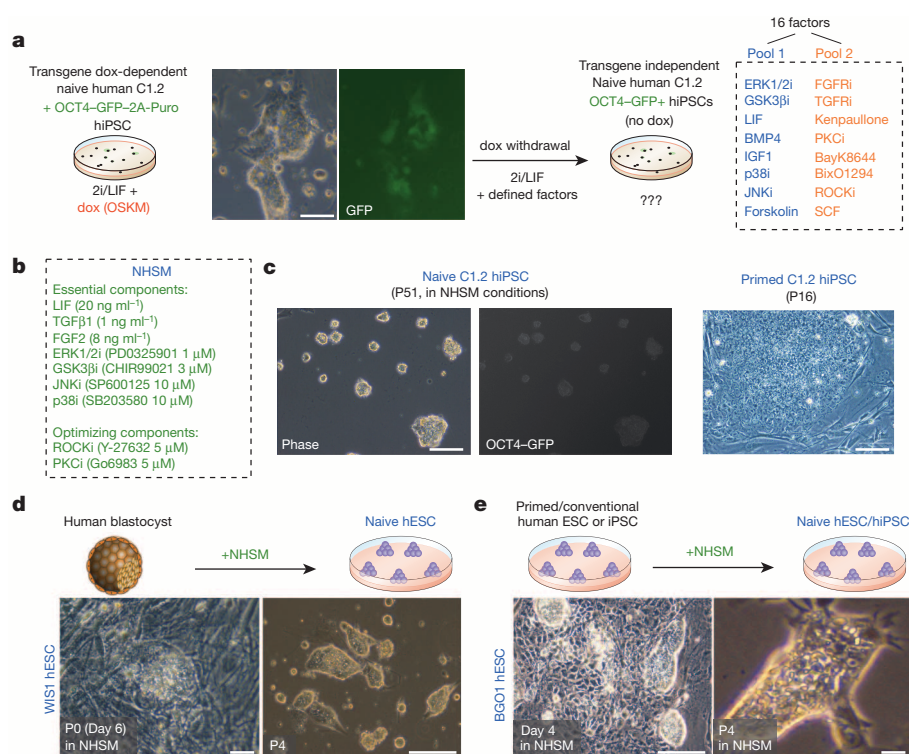


Figure 1 | Capturing the ground state of human naive pluripotency. **a**, Schematic drawing of the strategy used for calibrating conditions to isolate naive transgene-independent human iPS cells (hiPSCs) in the presence of 2i/LIF. Inhibitor is abbreviated as 'i' in the dotted box. Scale bar, 200 μm. **b**, Components of optimized NHSM conditions. **c**, Representative large-field view of OCT4-GFP⁺ human iPS cell colonies grown in NHSM (left) and, for comparison, primed/conventional iPS cells (right). Scale bars, left, 200 μm; right, 50 μm. **d**, Human-blastocyst-derived inner cell masses were plated on feeder cells in NHSM conditions. At day 6–8, the original outgrowth was trypsinized, and naive pluripotent cell lines were established in NHSM conditions (representative images of established line at the indicated passage (P) are shown). hESC, human embryonic stem cell. Scale bars, left, 50 μm; right, 200 μm. **e**, Epigenetic reversion of already established primed/conventional human ES cells to naive ground state. Representative images of BGO1 human ES cell line are shown. Scale bars, left, 200 μm; right, 30 μm.

in NHSM have domed-shaped colonies resembling murine naive cells; thus we refer to the selected cells as naive human pluripotent stem cells (hPSCs; includes human iPS cells and human ES cells), as we systematically validate the evidence for this claim.

We examined whether NHSM conditions allow derivation of new human ES cell lines from human blastocysts. Human-blastocyst-derived inner cell masses were plated in NHSM conditions, and successfully generated domed cell outgrowths after 6–8 days (Fig. 1d). We were able to establish four newly derived naive human ES cell lines termed LIS1, LIS2, WIS1 and WIS2 (Fig. 1d and Extended Data Fig. 1g). Several conventional (hereafter will be named 'primed') already established human ES cell lines (H1, H9, BGO1, WIBR1, WIBR2 and WIBR3) and human iPS cell lines (C1 and C2) were plated on gelatin/vitronectin-coated dishes in NHSM medium. After 4–8 days, dome-shaped colonies with packed round cell morphology could be readily isolated and further expanded (Fig. 1e). Human fibroblast cells were reprogrammed to human iPS cells in NHSM following reprogramming factor transduction (data not shown). All human ES cell and iPS cell lines expanded in NHSM conditions were positive for pluripotent markers and formed mature teratomas *in vivo* (Extended Data Fig. 2). Human naive pluripotent lines maintained normal karyotype after extended passaging following trypsinization and expansion on irradiated mouse embryonic fibroblast (MEF) feeder cells or on vitronectin/gelatin-coated plates (Extended Data Fig. 2a). The average doubling time was significantly reduced from 26 h for primed hPSCs down to 14 h for naive hPSCs (Extended Data Fig. 3a). Naive hPSCs displayed up to 35% single-cell cloning efficiency after trypsinization and sorting (without the use of ROCK inhibitors), whereas primed hPSCs largely did not survive single-cell cloning (Extended Data Fig. 3b). In the presence of ROCK inhibitor, naive hPSCs had single-cell cloning efficiency of up to 88%, whereas that of primed cells increased only up to 22% (Extended Data Fig. 3b, c).

We next aimed to characterize epigenetic features in naive human iPS cells and human ES cells established in NHSM conditions. Primed and naive human ES cell/human iPS cells were transfected with a luciferase reporter construct under the control of either the human distal

(DE) or the proximal (PE) enhancer sequences that control expression of the *OCT4* gene (also known as *POU5F1*). Consistent with previous reports, primed human ES cells showed preferential activation of the proximal *OCT4* enhancer element as typically seen in murine EpiSCs^{6,7}. Remarkably, predominant utilization of the *OCT4* distal enhancer was detected in naive human ES cells and human iPS cells (Extended Data Fig. 3d). To further substantiate these findings, WIBR3 human ES cells were stably transfected with full-length OCT4-GFP-2A-PURO, ΔPE-OCT4-GFP-2A-PURO or ΔDE-OCT4-GFP-2A-PURO engineered BAC reporter constructs (Extended Data Fig. 3e). The wild-type OCT4-GFP reporter was specifically active in both naive and primed conditions (Fig. 2a). The ΔPE-OCT4-GFP reporter was predominantly active in naive growth conditions whereas the ΔDE-OCT4-GFP reporter was more active in primed pluripotent cells (Fig. 2a). We then analysed the frequency and properties of X inactivation state¹² in several naive human ES cell/human iPS cell lines. Naive hPSCs captured in NHSM maintain a pre-X inactivation state as evident by nearly complete lack of H3K27me₃ nuclear foci and down regulation of *XIST* transcription. The majority of primed hPSCs demonstrated X inactivation as evident by the presence of H3K27me₃ nuclear foci and methylation of one of the *XIST* gene alleles (Fig. 2b and Extended Data Fig. 4a–d)¹³. Genome-wide mapping of H3K9me₃ by chromatin immunoprecipitation, followed by deep sequencing (ChIP-seq) in female and male naive and primed human pluripotent cells, indicated a significant increase ($P < 3.8 \times 10^{-63}$) in this mark on the X chromosome only in female primed pluripotent cells (Extended Data Fig. 4c). *XIST* promoter alleles are demethylated in male and female naive ES cells (Extended Data Fig. 4d). Upon differentiation of female naive human ES cells/human iPS cells, inactivation of one of the X chromosomes alleles becomes evident as the cells demonstrate H3K27me₃ clouds, upregulate *XIST* transcription simultaneously with methylation of one of the promoters of *XIST* alleles (Fig. 2b and Extended Data Fig. 4a–d).

Next we compared global gene expression patterns between naive and primed human ES cells and human iPS cells. Unbiased clustering of genome-wide expression profiles demonstrated that naive human

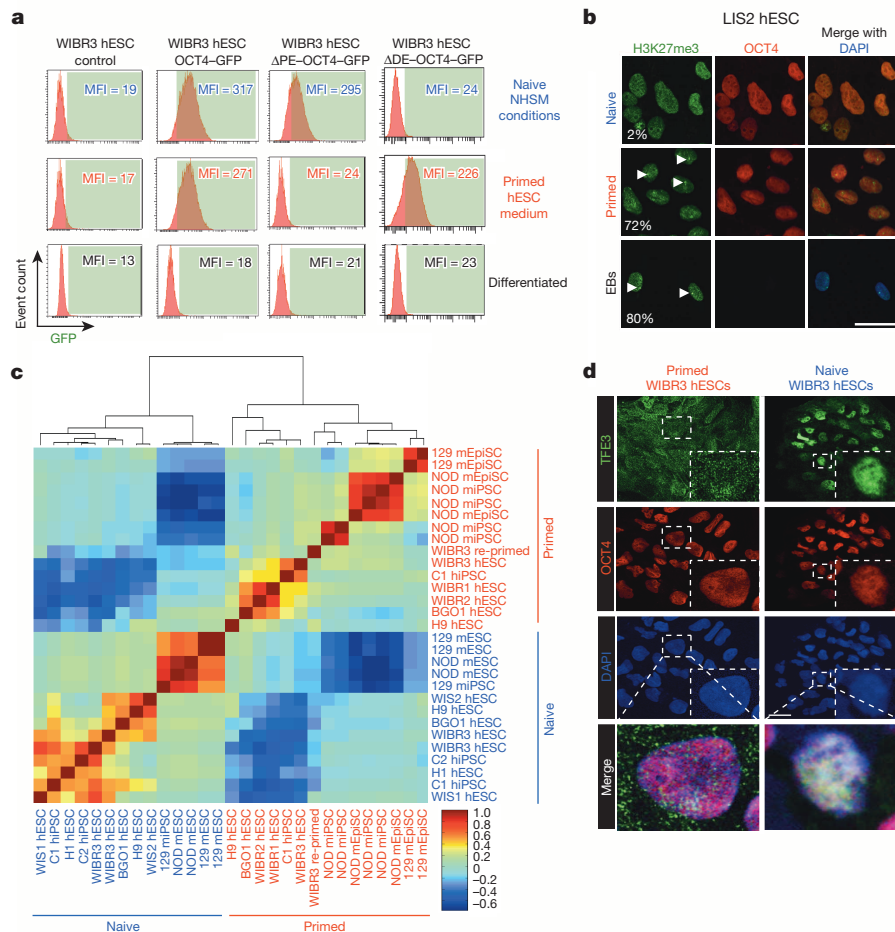


Figure 2 | Naive human stem cells share defining molecular features with mouse ES cells.

a, WIBR3 human ES cells were stably transfected with the different engineered BAC reporter constructs indicated. GFP levels were evaluated in genetically matched cells by flow cytometry. MFI, mean fluorescence intensity. Average MFI values of 3 biological replicates per sample are indicated. **b**, Representative confocal images obtained after double immunostaining for OCT4 and H3K27me3 on naive, primed and differentiated samples (obtained from naive cells). Average percentages of H3K27me3 foci-positive nuclei out of 150–200 cells per sample (arrowheads) are indicated. EBs, embryonic body differentiated cells. **c**, Unbiased global transcriptional cross-species hierarchical clustering of naive and primed pluripotent cells from mice and humans. Correlation matrix of gene expression was clustered using Spearman correlation and average linkage. Colour bar indicates correlation strength. Each row/column represents an independent cell line or clone. mEpiSCs, murine EpiSCs; mESC, mouse ES cell; miPSC, mouse iPS cell. NOD indicates non-obese diabetic ICR mice. **d**, Naive and primed human ES cells were double immunostained for TFE3 and OCT4. Representative confocal images are shown for WIBR3 human ES cells. Insets are enlargements of the dashed boxes. Scale bars, 50 μ m.

ES cells and human iPS cells possess a distinct gene expression pattern and clustered separately from primed human ES cells and human iPS cells (Extended Data Fig. 5a). Multiple transcripts associated with naive pluripotency were significantly upregulated in naive cells, including members of the *NANOG* and *DUSP* gene families (Extended Data Fig. 5b)¹⁴. Importantly, naive pluripotent cells had profound down-regulation of transcripts associated with lineage-commitment genes, including *ZIC1*, *SOX6* and *SOX11* that were expressed at low, but appreciable, levels in primed human ES cells (Extended Data Fig. 5b)³. Functional annotation analysis of differentially expressed genes with Gene Ontology (GO) revealed that genes downregulated in mouse and human naive pluripotency were significantly enriched for GO terms linked to developmental processes (Extended Data Fig. 5c). Furthermore, hierarchical clustering showed that human inner cell mass samples are transcriptionally more similar to naive human ES cells than to primed cells (Extended Data Fig. 5d). Primed human ES cells demonstrate intermediate expression levels of MHC class I surface antigen in comparison to somatic cells, whereas naive human ES cell/human iPS cell express only trace levels of MHC class I (Extended Data Fig. 6a)⁷. Moreover, although E-CADHERIN is expressed in primed human ES cells, the surface expression pattern becomes more prominent in naive human ES cell colonies (Extended Data Fig. 6b). Next we conducted an unbiased cross-species hierarchical clustering of the globally measured transcriptome to evaluate whether the human primed and naive pluripotent cells, described herein, globally correspond to those established in mice. By applying an algorithmic cross-species gene expression analysis on all 9,803 mouse–human orthologous genes found in our gene expression data sets^{7,8}, we found that whereas primed human ES cells and human iPS cells clustered with murine EpiSCs, all naive human ES cells/human iPS cells clustered with naive

mouse ES cells/mouse iPS cells independent of the genetic background or naive growth conditions used (Fig. 2c). Further, primed mouse and human pluripotent cells expression patterns were significantly more heterogeneous in comparison with their naive counterparts, as reflected by the noise distributions of the two groups in both species (Extended Data Fig. 5e). Finally, nuclear localization of the transcription factor TFE3 was recently shown to be enhanced in naive mouse ES cells, and compromised upon pluripotency priming¹⁵. Remarkably, a similar nuclear enrichment pattern for TFE3 was evident in naive human ES cells, and relative enrichment was compromised in primed human ES cells (Fig. 2d, Extended Data Fig. 6c). Collectively, our data demonstrates that NHSM endows human pluripotent cells with defining features of gene expression patterns typically seen in ground state naive mouse ES cells^{3,8}.

We moved on to map H3K4me3 and H3K27me3 chromatin marks by using ChIP-seq in mouse and human naive and primed pluripotent cells. Whereas distribution of both epigenetic markers over all genes promoters and bodies in mouse ES cells showed a significant ($P < 2 \times 10^{-37}$) decrease in naive conditions (Extended Data Fig. 4e), an even more notable decrease of both marks in mouse ES cells was observed over developmental genes ($P < 1.6 \times 10^{-71}$; Fig. 3a, right)³. Similarly, there was a marked ($P < 8.6 \times 10^{-61}$) reduction of H3K27me3 in promoter and gene-body region over developmental genes ($n = 5,922$) in human naive cells compared to primed (Fig. 3a, b). The reduction of H3K27me3 mark to nearly background levels over developmental genes in human naive cells was also reflected in the number of genes with bivalent marks near their TSS, which is more than 13-fold higher in primed (3,013 genes) compared to naive cells (226 genes). Notably, the naive pluripotency regulator, *TBX3*, acquired repressive H3K27me3 marks in both primed human and mouse samples (Extended Data

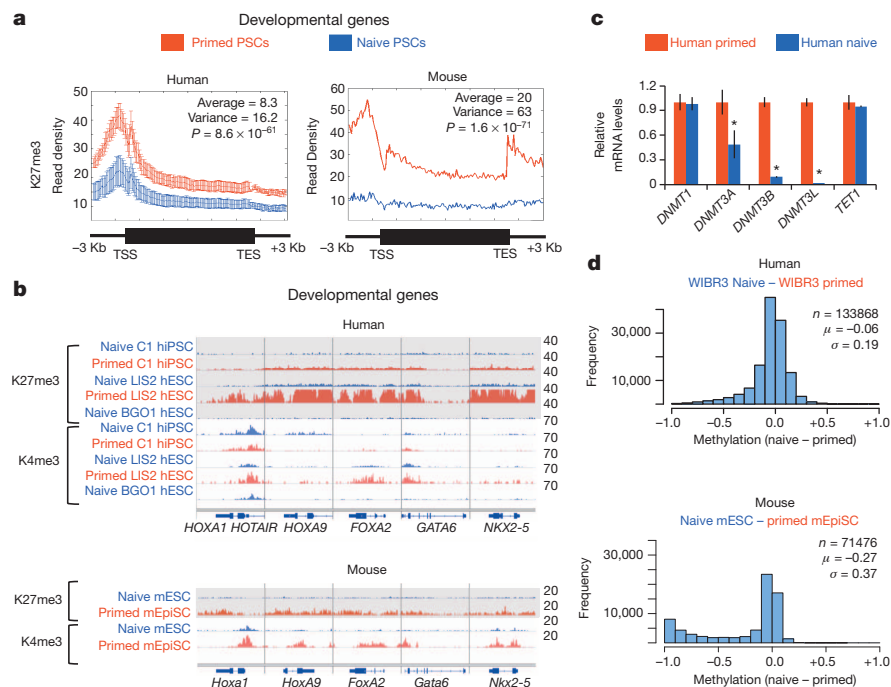


Figure 3 | Ground state epigenetic landscape of human naive pluripotency. **a**, Profiles of H3K27me3 chromatin mark over developmental genes in human and mouse, represented as normalized read-density. Blue, naive; red, primed. Human profiles indicate average and s.d. (error bars) calculated over 5 different cell lines (C1, LIS2, WIBR3, WIBR3-MBD3^{mut} and BGO1). Average difference between plots is indicated alongside variance and P -values (calculated with paired-sample t -test). **b**, Chromatin landscape of selected 5 developmental regulatory genes. H3K27me3 and H3K4me3 marks are shown for naive (blue) and primed (red) cell lines, in both human and mouse, showing high consistency between both species. **c**, Representative relative transcript levels in human primed and naive WIBR3 human ES cells after 48 h in NHSM conditions. * t -test P value < 0.01 . Error bars indicate s.d. ($n = 3$). **d**, Histograms of the change in methylation between primed and naive samples in human and mouse. The histograms depict the distribution of the per-CpG difference in methylation (naive methylation – primed methylation), calculated for all differentially methylated CpGs residing in CpG-rich regions ($>4\%$ CpG content) and having a coverage of $\geq \times 10$ in both analysed samples. n , the number of covered, differentially methylated CpGs. μ , σ , the mean and standard deviation of the distribution, respectively.

Fig. 7b). H3K27me3 peaks undergo genomic redistribution and become preferentially depleted from promoters and gene-body regions, rather than from intergenic regions (Extended Data Fig. 7a, b). We next measured H3K4me1 and H3K27ac marks using ChIP-seq, and globally mapped enhancers of class I and class II in naive and primed pluripotency states. Class I enhancers are characterized by the presence of H3K4me1 and H3K27ac, and the absence of H3K4me3 and H3K27me3 marks, and are associated with active expression of nearby genes¹⁶. Class II enhancers, on the other hand, are characterized by the presence of H3K4me1 and H3K27me3, and the absence of H3K27ac mark, and are associated with being poised for activation (Extended Data Fig. 7c, e)¹⁶. Consistent with the reduction of lineage-priming molecular features in naive pluripotency, we observed a major reduction in the annotated Class II enhancers in naive cell state, in both human (5.8-fold) and mouse (62.5-fold) (Extended Data Fig. 7c). Finally, reminiscent of recent reports on murine stem cells¹⁷, human naive ES cells/iPS cells expanded in NHSM conditions demonstrated a marked downregulation in DNMT3A, DNMT3B and DNMT3L *de novo* DNA methyltransferase enzymes, but not in DNMT1 and TET enzymes (Fig. 3c and Extended Data Fig. 6d). Sampling the DNA methylation states of human naive and primed pluripotent cells by reduced representation bisulphite sequencing (RRBS) showed a significant reduction (paired sample t -test P value ≈ 0) in CpG methylation in human naive WIBR3 ES cells after expansion in NHSM conditions for 17 days (Fig. 3d).

We next characterized and validated the distinct response patterns of naive and primed human pluripotent cells to different signalling stimuli or inhibitors, and compared them to those observed in rodent pluripotent cells (Extended Data Fig. 8a)⁷. Murine ES cells maintain their naive pluripotent state as measured by Δ PE–OCT4–GFP expression and by all-ES cell mouse formation competence following tetraploid embryo complementation assay, upon inhibition of either one of the p38, JNK or ERK pathways (Extended Data Fig. 8a–d). In human naive cells, simultaneous inhibition of all these pathways is essential for maintaining naive pluripotency (Extended Data Fig. 8a, f and Extended Data Fig. 6e). Supplementation of TGF- β is necessary to inhibit the pro-neural differentiation effect of ERK inhibition in naive NHSM conditions (Extended Data Fig. 6f). Naive human ES cells are LIF/STAT3-dependent and when stably transfected with dominant-negative

Stat3-encoding transgenes, they rapidly differentiated and could not be maintained *in vitro* with NHSM conditions (Extended Data Fig. 8a, g). Alternatively, human naive cells transgenic for a constitutively active Stat3 mutant (Stat3-CA) could be propagated *in vitro* in the absence of exogenous LIF (Extended Data Fig. 8g). These findings highlight that several signalling pathways distinctly regulate the two human pluripotent states, in a manner similar to that observed in mice^{3,7}.

We next evaluated whether the altered biological features of human naive PSCs directly endow them with unique or enhanced functional characteristics. Naive H9 human ES cells were significantly more amenable to gene targeting by homologous recombination with isogenic vectors, in comparison to H9-primed cells (Extended Data Fig. 9a, b). NHSM conditions were important for facilitating up to 100% efficiency in deriving naive human iPS cell formation from MBD3-depleted *in vitro*-differentiated secondary OKSM transgenic fibroblasts following dox induction (Extended Data Fig. 9c)¹⁸. Consistent with changes in DNA methylation regulation in naive human pluripotency (Fig. 3d), iPS cells from fragile X chromosome patient specific fibroblasts were able to undergo *FMR1* gene reactivation and demethylation after reprogramming under naive ground state conditions (Extended Data Fig. 9d, e)¹⁹. Finally, consistent with robust E-CADHERIN expression, enhanced single-cell survival and corresponding to an earlier developmental state, microinjection of GFP-labelled human naive iPS cells into mouse E2.5 morulas showed robust integration in mouse *ex vivo*-developed inner cell masses at E3.5, and higher integration than that of primed pluripotent cells (Extended Data Fig. 10a–c). The latter results are consistent with previously reported near-absence of integration and chimaerism following embryo micro-injection of primate primed ES cells^{20,21}. Mouse blastocysts obtained after microinjection with human naive iPS cells were implanted in pseudo-pregnant female mice and allowed to develop for 7–8 additional days *in vivo*, after which they were dissected and imaged *in toto* under confocal microscopy for GFP⁺ cell detection. Remarkably, we were able to obtain multiple ($n = 13$) mouse embryos corresponding to E8.5–E10.5 developmental stages that showed chimaerism with naive human iPS-cell-derived GFP⁺ cells integrated in organogenesis stages of embryonic development and in different locations, including craniofacial tissues and embryonic neural folds (representative examples in Fig. 4, Extended Data Fig. 10d, e and Supplementary Videos

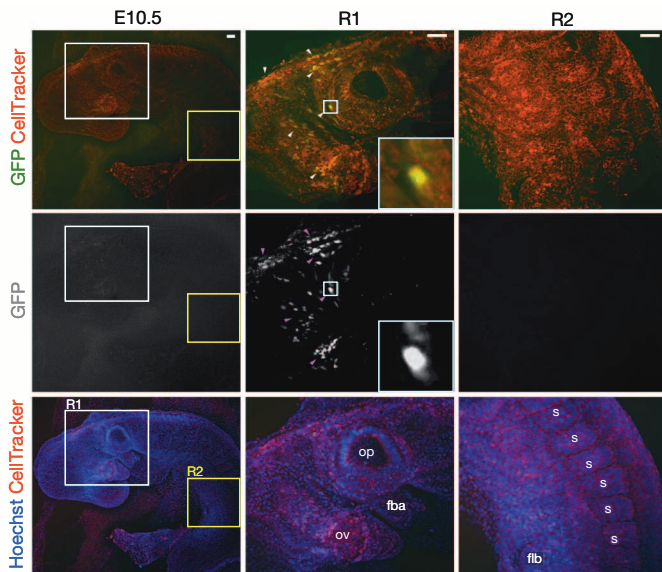


Figure 4 | Robust generation of cross-species chimaeric humanized mice following naive human iPS cell microinjection into mouse morulas.

Representative images showing widespread integration of GFP-labelled human naive iPS-derived cells into different locations in the anterior part of an E10.5 mouse embryo. Hoechst and CellTracker were used for counterstaining. The first column shows the whole embryo (z-stack interval 30 μ m, 18 focal planes in total). The second column shows a zoom in images focusing on the head region (white square R1) where the human iPS-derived cells (GFP-positive cells) are pointed out (arrowheads, z-stack interval 20 μ m, 11 focal planes in total). The third column shows the posterior part of the embryo (yellow square R2) where no GFP-positive cells were detected (z-stack interval 20 μ m, 9 focal planes in total). Squares in the first two images in the second column represent the area shown in the insets at the corner of each image. ov, optic vesicle; op, otic pit; fba, first branchial arch; flb, forelimb bud; s, somite. Scale bars, 50 μ m.

1 and 2). In-depth functional and developmental analysis of the *in vivo*-integrating human naive iPS-cell-derived cells in advanced cross-species chimaeric mouse embryos is of great future scientific interest.

Our findings substantiate the concept of naive ground state pluripotency in humans^{7,8}, and indicate that its maintenance requires a unique combination of cytokines and small molecule inhibitors. The epigenetic changes induced by NHSM conditions (Fig. 3) indicate that naive conditions are likely to resolve previously described technical phenotypes of epigenetic memory, lineage differentiation biases and aberrant reprogramming in human iPSCs and ES cells¹. Finally, defining a novel naive pluripotent state in humans that is stable and requires no genetic modifications might be relevant for the molecular study of early lineage commitment, and for expanding the capabilities for using human iPS cells in regenerative medicine research and disease modelling *in vitro* and *in vivo* through the generation of 'humanized' animal models.

METHODS SUMMARY

Full details of stem cell lines, cell culture, plasmids, antibodies, embryo-micro-manipulation and bioinformatics are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 May; accepted 10 October 2013.

Published online 30 October; corrected online 11 December 2013 (see full-text HTML version for details).

1. Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508–525 (2010).

2. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
3. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
4. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
5. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
6. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
7. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl Acad. Sci. USA* **107**, 9222–9227 (2010).
8. Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513–524 (2009).
9. Greber, B. *et al.* Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6**, 215–226 (2010).
10. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnol.* **25**, 681–686 (2007).
11. Rajendran, G. *et al.* Inhibition of protein kinase C signaling maintains rat embryonic stem cell pluripotency. *J. Biol. Chem.* **288**, 24351–24362 (2013).
12. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
13. Mekhoubad, S. *et al.* Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012).
14. Chappell, J., Sun, Y., Singh, A. & Dalton, S. MYC/MAX control ERK signaling and pluripotency by regulation of dual-specificity phosphatases 2 and 7. *Genes Dev.* **27**, 725–733 (2013).
15. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335–347 (2013).
16. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
17. Fic, G. *et al.* Fgf signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
18. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
19. Urbach, A., Bar-Nur, O., Daley, G. Q. & Benvenisty, N. Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells. *Cell Stem Cell* **6**, 407–411 (2010).
20. James, D., Noggle, S. A., Swigut, T. & Brivanlou, A. H. Contribution of human embryonic stem cells to mouse blastocysts. *Dev. Biol.* **295**, 90–102 (2006).
21. Tachibana, M., Sparman, M., Ramsey, C., Ma, H. & Lee, H. S. Generation of chimeric rhesus monkeys. *Cell* **148**, 285–295 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements J.H.H. is supported by a gift from Ilana and Pascal Mantoux, and research grants from the European Research Council starting investigator grant (StG-2011-281906, for funding iPS cell experiments only), the Leona M. and Harry B. Helmsley Charitable Trust, the BIRAX (Britain Israel Research and Academic Exchange Partnership), The Sir Charles Clore Research Prize, the Israel Science Foundation (Bikura, ICORE (Israeli Centre of Research Excellence) and Regular research program), the Israel Cancer Research Foundation, the ERANET E-Rare disease program, the Benozio Endowment fund, Fritz Thyssen Stiftung (used for human iPS cell experiments only), EMBO young investigator program, the Alon Foundation scholar award, a grant from E. A. and R. Drake, postdoctoral fellowships from ICRF and the Weizmann Dean fellowship award for A.A.M. We thank the embryologists of the Racine In Vitro Fertilization Laboratory at the Lis Maternity Hospital (A. Carmon, T. Cohen and N. M. Raz) for their skilful assistance. We thank the Weizmann Institute management for providing financial and infrastructural support.

Author Contributions J.H.H. conceived the idea for this project, designed and conducted experiments, invented and optimized NHSM conditions, and wrote the manuscript with contributions from other authors. L.W. and A.A.M. conducted tissue culture optimization and molecular biology analysis. O.G. and M.Z. conducted microinjections. O.G. and R.M. conducted embryo imaging. E.C., Z.S., Z.M. and A.T. conducted RRBS analysis. Y.S.M. analysed gene expression data. N.N. and E.C. analyzed ChIP-seq data. J.H.H., Y.K., T.S. and D.B.-Y. applied NHSM conditions on human blastocysts. O.G., I.M., I.A., S.Gi., D.A.-Z. and S.B. assisted in ChIP-seq experiments. S.V. and J.H.H. engineered human stem cell lines. Y.R., V.K., I.C., D.S., S.Ge. assisted and participated in iPS cell related experimental assays. A.Z. conducted automated imaging analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.H. (jacob.hanna@weizmann.ac.il), N.N. (noa.novershtern@weizmann.ac.il) or R.M. (rada.massarwa@weizmann.ac.il).

A chain mechanism for flagellum growth

Lewis D. B. Evans¹, Simon Poulter¹, Eugene M. Terentjev², Colin Hughes¹ & Gillian M. Fraser¹

Bacteria swim by means of long flagella extending from the cell surface. These are assembled from thousands of protein subunits translocated across the cell membrane by an export machinery at the base of each flagellum. Unfolded subunits^{1–3} then transit through a narrow channel at the core of the growing flagellum to the tip, where they crystallize into the nascent structure. As the flagellum lengthens outside the cell, the rate of flagellum growth does not change⁴. The mystery is how subunit transit is maintained at a constant rate without a discernible energy source in the channel of the external flagellum⁵. We present evidence for a simple physical mechanism for flagellum growth that harnesses the entropic force of the unfolded subunits themselves. We show that a subunit docked at the export machinery can be captured by a free subunit through head-to-tail linkage of juxtaposed amino (N)- and carboxy (C)-terminal helices. We propose that sequential rounds of linkage would generate a multisubunit chain that pulls successive subunits into and through the channel to the flagellum tip, and by isolating filaments growing on bacterial cells we reveal the predicted chain of head-to-tail linked subunits in the transit channel of flagella. Thermodynamic analysis confirms that links in the subunit chain can withstand the pulling force generated by rounds of subunit crystallization at the flagellum tip, and polymer theory predicts that as the N terminus of each unfolded subunit crystallizes, the entropic force at the subunit C terminus would increase, rapidly overcoming the threshold required to pull the next subunit from the export machinery. This pulling force would adjust automatically over the increasing length of the growing flagellum, maintaining a constant rate of subunit delivery to the tip.

The structural subunits of the sequential rod, hook and flagellin filament of the bacterial flagellum (Extended Data Fig. 1) are unfolded and translocated across the cell membrane by a type III export machinery^{2,6}, energised by ATP hydrolysis^{1,7} and the proton motive force^{5,8,9}. They pass into a central channel that extends the length of the growing flagellum outside the cell^{10–12}. The channel has a diameter of around 20 Å, so subunits must remain unfolded as they transit to the flagellum tip up to 15–20 µm, a distance that is 10 cell lengths away², where they crystallize into the growing structure¹³. Theoretical models describing the physical challenges presented by assembling flagella outside the living cell are set out in Supplementary Information section 1, but how this feat is achieved remains a mystery. Models based on pushing or pumping mechanisms^{14,15} operating in the cell at the membrane entrance to the flagellum channel, for example, powered by the proton motive force, appear incompatible with constant rate transit, as lengthening of the pushed subunit column would engender a proportional increase in resistance and slowing of growth. Furthermore, the intrinsic flexibility of unfolded subunits would preclude transmission of compressive forces¹⁶. An alternative hypothesis¹⁷, involving diffusion of flagellin, relies on subunits adopting an unprecedented extended α -helical fold, and does not take into account interactions between subunits in the channel. In contrast to passive diffusion, the notion of active transport of subunits through the channel by a Brownian ratchet¹⁸ is called into question by the absence of ATP in the channel and the channel being too narrow to accommodate hypothetical motors.

We set out to examine experimentally a fundamentally distinct hypothesis, in which we propose that the energy for subunit transit though

the channel resides in the unfolded subunits themselves as they move from the export machinery at the base of the flagellum^{2,19,20}. We reasoned that such movement might be achieved by capture of each docked subunit by the preceding subunit already in the channel. Direct support for this possibility is provided by challenge experiments (Fig. 1) in which free flagellar hook subunit FlgE effected concentration-dependent release of another subunit docked at the export machinery gate component (that is, in preformed subunit-gate complexes, FlgD–[GST–FlhB_C]; FlhB_C stands for the C-terminal cytoplasmic domain of FlhB). Subunits docked at the gate component were captured, not displaced, by free subunit as the challenge subunit FlgE was engineered such that it was unable to bind the export gate (gate-blind; Extended Data Figs 2 and 3), and heteromeric FlgD–FlgE subunit–subunit capture complexes were confirmed by affinity chromatography (Fig. 1c). Further capture assays established that gate-docked FlgD can be captured by other hook subunits, but not by the filament cap subunit that is assembled later^{21,22} (Extended Data Fig. 4). We supposed that in the subunit docked at the export machinery, the extreme N terminus would be available to the free C terminus of the preceding subunit already in the channel, and that linkage of these juxtaposed termini could effect capture. Evidence for this was obtained by showing that C-terminally truncated gate-blind FlgE challenge subunit was unable to release gate-docked subunit in a concentration-dependent manner from the FlgD–FlhB_C subunit-gate complex (Fig. 1b). We reasoned that links between N and C termini of unfolded adjacent subunits might adopt a parallel coiled-coil conformation (Fig. 1d) and using unique cysteine–cysteine crosslinks (FlgE–Ct V₄₀₀C with FlgE–Nt A₁₄C, D₁₈C or A₂₅C), we trapped a subunit dimer compatible with just such a head-to-tail arrangement (Fig. 1e). Weaker crosslinking was detected between FlgE–Ct V₄₀₀C and FlgE–Nt A₆C, which are at opposite ends of the predicted coiled-coil, and no complexes were detected for FlgE–Nt A₄₀C, which lies out with the subunit extreme N-terminal helix (Fig. 1e). This conformation is different from that observed in assembled flagella, in which subunit termini²³ fold as antiparallel coiled-coils that line the channel^{10,12} (Extended Data Fig. 5). The 8 Å cysteine–cysteine crosslinks could not form between subunits in the assembled hook, as the cysteines are too far apart^{12,24} (Extended Data Table 1). The crosslinking data show that subunit capture is achieved by subunits linking head-to-tail through their terminal helices, predicted to form a parallel coiled-coil, and indicate that each subunit terminus contributes 14–25 residues to each coiled-coil link.

We suggest that successive rounds of such linkage could generate a chain of subunits that extends through the flagellar channel. However, although a single unfolded hook subunit could span the channel in the rod/hook structure (~90 nm), crystallizing into the tip of the hook as it simultaneously recruits the next subunit from the export machinery, the flagellar filament grows up to 20 µm (ref. 2). Our proposed chain mechanism would therefore require many flagellin (FliC) subunits to link in the same way to span the transit channel of the flagellar filament. We therefore set out to visualize the linked flagellin subunits *in vivo* by isolating part of the predicted flagellin chain from within the flagellar filament growing on the surface of living cells. First, again using *in vitro* cysteine–cysteine crosslinking, we confirmed the same type of head-to-tail linkage between flagellin subunits that we had shown between hook subunits (Fig. 2a, FliC–CtQ₄₈₈C or N₄₈₉C with FliC–Nt S₁₁C, L₁₃C,

¹Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK. ²Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK.

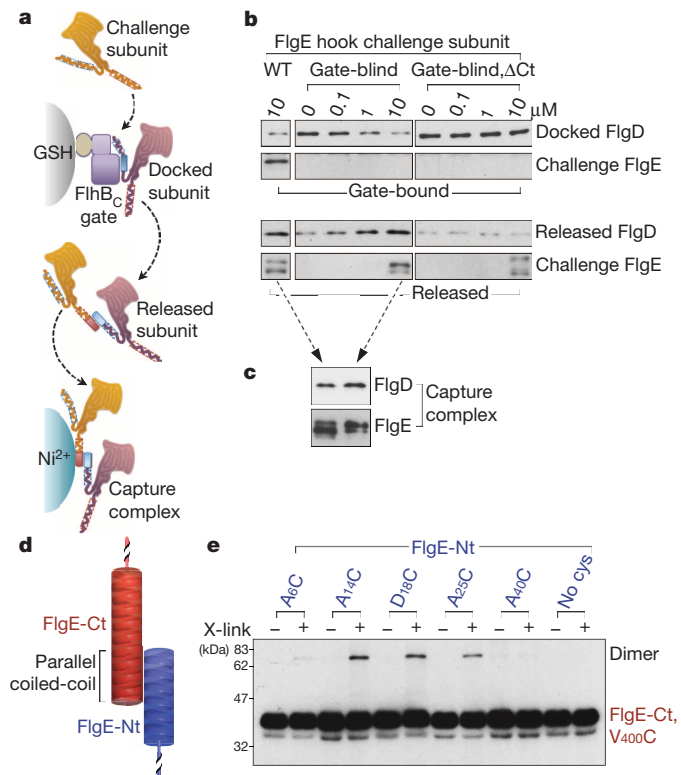


Figure 1 | Free subunits capture subunits docked at the export gate through head-to-tail linkage of terminal helices. **a**, Capture of FlgD subunits docked at glutathione (GSH) bead-immobilized GST-FlhB_C export gate component by free challenge subunits. Released capture complexes were collected by Ni²⁺ affinity chromatography. **b**, Docked FlgD subunit was challenged with increasing concentrations of free FlgE hook subunit, either wild type (WT) or unable to bind the export gate (gate-blind; Extended Data Fig. 3) or gate-blind and lacking C-terminal residues 359–403 (gate-blind, ΔCt). FlgE subunit lacking the C terminus was attenuated for export (Extended Data Fig. 4e). **c**, Released capture complexes generated in **b**, (dashed arrows) were isolated confirming that captured FlgD is linked to challenge subunit. **d**, Juxtaposed carboxyl (FlgE-Ct)- and amino (FlgE-Nt)-terminal helices of sequential subunits adopt a parallel coiled-coil. **e**, Subunit pairs linked head-to-tail (dimer) using *in vitro* site-specific cysteine–cysteine crosslinks. FlgE and variants containing unique cysteines (A₆C, A₁₄C, D₁₈C, A₂₅C or A₄₀C; Extended Data Table 2) within and adjacent to the N-terminal helix predicted to generate the coiled-coil were incubated, with (+) or without (–) BMOE cross-linker (x-link) and a FlgE variant (V₄₀₀C) containing a unique cysteine within the C-terminal helix. FlgE derivatives lacked either C or N termini (FlgE-Nt, FlgE-Ct) to preclude self-interaction. No dimers were detected for subunits without engineered cysteines (Extended Data Fig. 5e). All experiments were carried out at least three times and were biological replicates.

L₁₈C or R₃₁C, but not with FliC-Nt K₁₇₈C that lies out with the N-terminal helix; Extended Data Table 1). The crosslinking distances indicate that each flagellin terminus could contribute 21–32 residues to the coiled-coil link. We then set out to trap the predicted unfolded subunit chain during transit through the channel as the flagellum grows on bacteria. Using *in vivo* site-specific crosslinking of engineered flagellin containing cysteines shown to link *in vitro* (R₃₁C, Q₄₈₈C; Fig. 2a), we isolated higher order flagellin complexes and visualized, as a ladder, oligomeric chains containing three or more subunits (Fig. 2b). These oligomers could only form by head-to-tail linkage as a chain in the channel, and no chain was detected with flagellin containing R₃₁C alone, or K₁₇₈C and Q₄₈₈C, which cannot engender head-to-tail links. Our data show that there are multiple linked subunits in the channel, incompatible with the recent theoretical model of filament growth¹⁷ in which transiting subunits do not interact with each other in the channel.

Having provided direct evidence of subunit capture, linkage and the resulting chain, we propose a physical mechanism for subunit transit

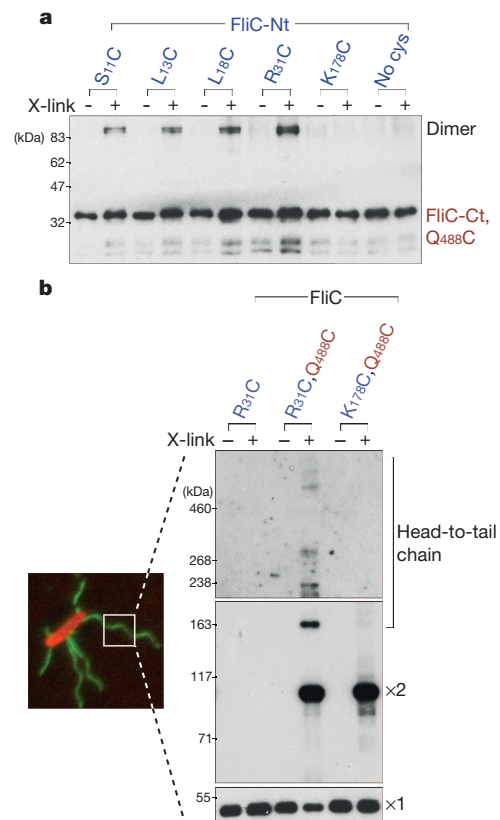


Figure 2 | Head-to-tail linkage of flagellin subunits assembles a chain in the flagellum growing on the bacterial cell surface. **a**, Flagellin pairs linked head-to-tail (dimer) using *in vitro* site-specific cysteine–cysteine crosslinks. Flagellin (FliC-Nt, FliC-Ct) and its variants containing unique cysteines (S₁₁C, L₁₃C, L₁₈C, R₃₁C or K₁₇₈C) within and adjacent to the N-terminal helix predicted to generate a coiled-coil were incubated, with (+) or without (–) BMOE cross-linker (x-link) and a FliC variant containing a unique cysteine (Q₄₈₈C) within the C-terminal helix. No complexes were detected between subunits without engineered cysteines (Extended Data Fig. 6a). **b**, Trapping of flagellin linked head-to-tail in chains within flagella (labelled green) growing on *Salmonella enterica* serovar Typhimurium (hereafter referred to as *S. typhimurium*) cells (labelled red) using *in vivo* site-specific cysteine–cysteine crosslinking. Cells expressing recombinant full-length flagellin containing engineered cysteines (right panels; left lane, negative control R₃₁C; centre lane, R₃₁C and Q₄₈₈C predicted to trap chain; right lane K₁₇₈C and Q₄₈₈C negative control) were incubated with (+) or without (–) BMOE cross-linker (x-link). Flagellar filaments were then isolated, depolymerized and resolved (immunoblot, panel exposure times decrease from top to bottom) to reveal monomer (×1), dimer (×2) and higher order head-to-tail chains of flagellin. We estimate between 6–10 linked subunits in the largest species. Dimers are predicted to also form by crosslinking between assembled flagellin C-terminal cysteines (Q₄₈₈C) that line the channel and subunits in transit, so dimers are also seen in the controls (Extended Data Fig. 6b, c). All experiments were carried out at least three times and were biological replicates.

energised by the thermal motion of the chain of linked unfolded subunits anchored at the flagellum tip. In this mechanism, depicted in Fig. 3 and supported with calculations in Supplementary Information section 2, subunit folding at the tip would gradually decrease the number of free residues remaining in the channel and intrinsically provide an increasing pulling force on the next linked subunit docked at the export machinery. The proposed mechanism requires strong anchoring of the subunit at the tip²⁵, and breaking this anchor would require an estimated a force F_A of ~11 nN. Strong anchoring (F_A of ~6–6.7 nN) could also be achieved in the absence of subunit assembly into the crystallized structure, explaining how some flagellar proteins, for example, filament cap and hook-filament junction proteins, are exported even when no longer required for flagellum assembly²⁶ (similar export without assembly is observed

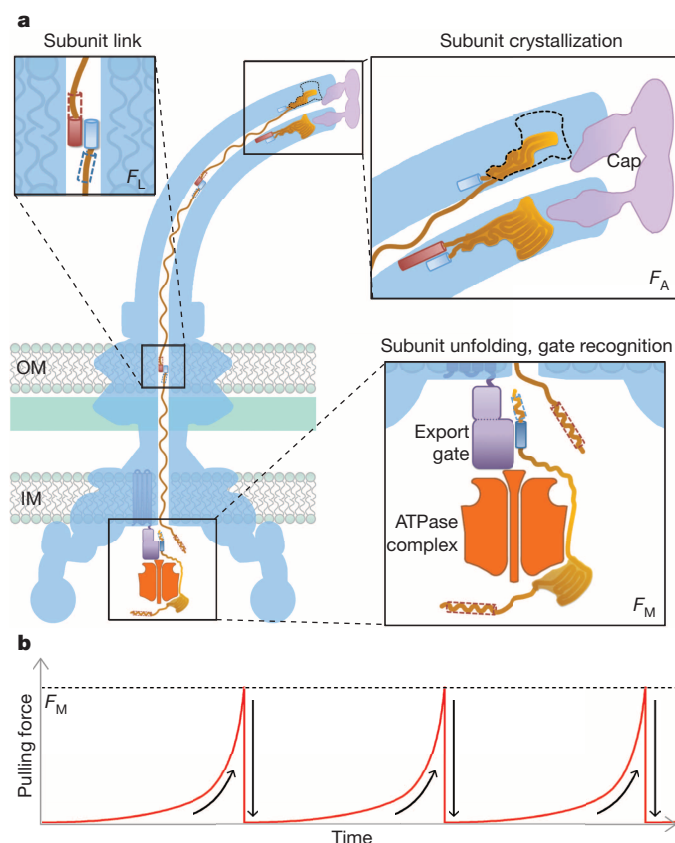


Figure 3 | An entropic chain mechanism for flagellum growth outside the cell. **a**, Subunit crystallization beneath the cap foldase¹³ provides a strong anchor (force to break anchor, F_A ; Supplementary Information section 2). Sequential subunits are linked (force to break link, F_L) head-to-tail in a chain by juxtaposed terminal helices forming parallel coiled-coils. Subunits docked at the export ATPase⁶ are unfolded¹ and docked by the export machinery. The N-terminal helix of the docked subunit is then captured (force to break docking, F_M) into the subunit chain by the free C-terminal helix of an exiting subunit in the flagellar channel. **b**, Successive rounds of subunit capture from the export machinery. As an unfolded subunit crystallizes, the pulling force at its free end increases to reach the F_M threshold (dashed line), whereupon the next subunit is captured from the export machinery. The pulling force then drops rapidly as the new unfolded subunit enters the channel. This process repeats for each subunit captured into the chain.

for effectors secreted through related virulence needles; Supplementary Information sections 2a and 3). Second, the mechanism requires parallel coiled-coil linking of sequential subunit termini and, as each hook subunit contributes 14–25 residues to the coiled-coil, breaking this link would require an estimated force F_L of 400–700 pN. Breaking the links between the more abundant flagellin subunits would require a force of the same order of magnitude. The chain mechanism requires that subunit anchoring at the flagellum distal tip must be stronger than subunit–subunit linking in the channel, which in turn must be stronger than the force F_M required to pull subunits from the export machinery into the channel. Affinity of subunit for the export gate component, measured by isothermal titration calorimetry (ITC) (Extended Data Fig. 2a), estimates F_M at ~ 30 pN, fulfilling the requirement $F_M < F_L < F_A$.

Polymer theory describes the entropic pulling force at the C terminus of the unfolded subunit chain anchored at its N terminus by crystallization into the flagellum, and explains how this force adjusts as channel length increases^{27,28} (Supplementary Information section 2c). In an anchored chain of N residues each of length a , when a distance R separates the export machinery from the anchor point at the distal tip, the force depends on R/Na (Extended Data Fig. 6). As a subunit crystallizes, fewer residues are left free in the channel (Na decreases), the pulling

force $F(R,Na)$ at the export machinery increases, eventually reaching the critical value F_M and the next subunit is pulled from the export machinery into the channel. Once this occurs, Na increases, the entropic force drops rapidly and the ‘slack’ of residues becomes available for a new round of subunit crystallization at the tip. This process repeats for each subunit pulled into the channel (Fig. 3b). The same principle applies even if the subunit chain breaks inside the channel, either spontaneously or by shearing of the flagellum. The most cell-distal subunit in the remaining chain could diffuse to the tip⁴ and start crystallizing, thus providing an increasing anchoring that would gradually restart the pulling mechanism. Crucially, the entropic force $F(R,Na)$ automatically adjusts with the length of the multi-subunit chain (Supplementary Information section 3 and Extended Data Table 2) engendering a constant rate of subunit transit independent of channel length. The chain mechanism we propose is therefore supported by both experimentation and thermodynamic analysis, and we suggest that it explains the published observations on how flagella grow outside the bacterial cell.

METHODS SUMMARY

Salmonella enterica serovar Typhimurium (*S. typhimurium*) strains and plasmids used in this study are described in Extended Data Table 3. Bacteria were cultured at 30–37 °C in Luria-Bertani broth containing, where appropriate, antibiotics and chemical inducers. Protein–protein interactions were assessed by ITC and affinity chromatography using purified proteins or lysates of cells expressing recombinant proteins. Interaction surfaces in protein complexes were identified by photo-crosslinking of unnatural amino acids or chemical crosslinking of site-specific engineered cysteines. Covalently linked complexes from crosslinking experiments exhibited a higher apparent molecular mass than the free proteins when samples were separated by SDS–PAGE. Capture assays were performed by challenging preformed subunit FlhB_C gate complexes with increasing concentrations of free subunit, and released subunit was subsequently affinity purified to assess the stability of capture complexes. *In vivo* flagellar subunit export assays were performed at mid-log growth in fresh warm media for 40 min, as previously described⁶. Proteins separated by SDS–PAGE were visualized by either Coomassie staining or immunoblotting using specific antibodies, either to an epitope tag or the recombinant protein. All experiments were carried out at least three times and were biological replicates.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 December 2012; accepted 23 September 2013.

Published online 10 November 2013.

1. Akeida, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, 911–915 (2005).
2. Chevance, F. F. & Hughes, K. T. Coordinating assembly of a bacterial macromolecular machine. *Nature Rev. Microbiol.* **6**, 455–465 (2008).
3. Abrusci, P. *et al.* Architecture of the major component of the type III secretion system export apparatus. *Nature Struct. Mol. Biol.* **20**, 99–104 (2013).
4. Turner, L., Stern, A. S. & Berg, H. C. Growth of flagellar filaments of *Escherichia coli* is independent of filament length. *J. Bacteriol.* **194**, 2437–2442 (2012).
5. Galán, J. E. Energizing type III secretion machines: what is the fuel? *Nature Struct. Mol. Biol.* **15**, 127–128 (2008).
6. Thomas, J., Stafford, G. P. & Hughes, C. Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. *Proc. Natl Acad. Sci. USA* **101**, 3945–3950 (2004).
7. Ibuki, T. *et al.* Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nature Struct. Mol. Biol.* **18**, 277–282 (2011).
8. Minamino, T. & Namba, K. Distinct roles of the Flil ATPase and proton motive force in bacterial flagellar protein export. *Nature* **451**, 485–488 (2008).
9. Paul, K., Erhardt, M., Hirano, T., Blair, D. F. & Hughes, K. T. Energy source of flagellar type III secretion. *Nature* **451**, 489–492 (2008).
10. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650 (2003).
11. Samatey, F. A. *et al.* Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062–1068 (2004).
12. Fujii, T., Kato, T. & Namba, K. Specific arrangement of alpha-helical coiled coils in the core domain of the bacterial flagellar hook for the universal joint function. *Structure* **17**, 1485–1493 (2009).
13. Yonekura, K. *et al.* The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science* **290**, 2148–2152 (2000).
14. Minamino, T., Imada, K. & Namba, K. Mechanisms of type III protein export for bacterial flagellar assembly. *Mol. Biosyst.* **4**, 1105–1115 (2008).

15. Tanner, D. E., Ma, W., Chen, Z. & Schulten, K. Theoretical and computational investigation of flagellin translocation and bacterial flagellum growth. *Biophys. J.* **100**, 2548–2556 (2011).
16. Blundell, J. R. & Terentjev, E. M. Buckling of semiflexible filaments under compression. *Soft Matter* **5**, 4015–4020 (2009).
17. Stern, A. & Berg, H. Single-file diffusion of flagellin in flagellar filaments. *Biophys. J.* **105**, 182–184 (2013).
18. Jülicher, F., Ajdari, A. & Prost, J. Modeling molecular motors. *Rev. Mod. Phys.* **69**, 1269–1282 (1997).
19. Minamino, T. & Macnab, R. M. Domain structure of *Salmonella* FlhB, a flagellar export component responsible for substrate specificity switching. *J. Bacteriol.* **182**, 4906–4914 (2000).
20. Evans, L. D., Stafford, G. P., Ahmed, S., Fraser, G. M. & Hughes, C. An escort mechanism for cycling of export chaperones during flagellum assembly. *Proc. Natl Acad. Sci. USA* **103**, 17474–17479 (2006).
21. Minamino, T., Doi, H. & Kutsukake, K. Substrate specificity switching of the flagellum-specific export apparatus during flagellar morphogenesis in *Salmonella typhimurium*. *Biosci. Biotechnol. Biochem.* **63**, 1301–1303 (1999).
22. Stafford, G. P. *et al.* Sorting of early and late flagellar subunits after docking at the membrane ATPase of the type III export pathway. *J. Mol. Biol.* **374**, 877–882 (2007).
23. Vonderviszt, F., Ishima, R., Akasaka, K. & Aizawa, S. Terminal disorder: a common structural feature of the axial proteins of bacterial flagellum? *J. Mol. Biol.* **226**, 575–579 (1992).
24. Shaikh, T. R. *et al.* A partial atomic structure for the flagellar hook of *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **102**, 1023–1028 (2005).
25. Vonderviszt, F., Závodszky, P., Ishimura, M., Uedaira, H. & Namba, K. Structural organization and assembly of flagellar hook protein from *Salmonella typhimurium*. *J. Mol. Biol.* **251**, 520–532 (1995).
26. Homma, M., Fujita, H., Yamaguchi, S. & Iino, T. Excretion of unassembled flagellin by *Salmonella typhimurium* mutants deficient in hook-associated proteins. *J. Bacteriol.* **159**, 1056–1059 (1984).
27. Sotta, P., Lesne, A. & Victor, J. M. Monte Carlo simulation of a grafted polymer chain confined in a tube. *J. Chem. Phys.* **112**, 1565–1573 (2000).
28. Blundell, J. R. & Terentjev, E. M. Semiflexible filaments subject to arbitrary interactions: a Metropolis Monte Carlo approach. *Soft Matter* **7**, 3967–3974 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Hinchliffe for help with analysis of atomic structures, K. Yonekura and K. Namba for providing information on filament structure, H. Berg and N. Greene for discussions, A. Crow for advice on ITC, and P. Dhillon for technical assistance. The work was supported by a Wellcome Trust Programme Grant (C.H. and G.M.F.).

Author Contributions L.D.B.E., S.P., C.H. and G.M.F. designed the experiments; L.D.B.E. and S.P. performed experiments; L.D.B.E., C.H. and G.M.F. analysed the data; E.M.T. performed the thermodynamic and polymer theory analyses; L.D.B.E., E.M.T., C.H. and G.M.F. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.M.F. (gmf25@cam.ac.uk).

High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy

Samuel A. Hasson^{1,2*}, Lesley A. Kane^{1*}, Koji Yamano¹, Chiu-Hui Huang¹, Danielle A. Sliter¹, Eugen Buehler², Chunxin Wang¹, Sabrina M. Heman-Ackah³, Tara Hessa¹, Rajarshi Guha², Scott E. Martin² & Richard J. Youle¹

An increasing body of evidence points to mitochondrial dysfunction as a contributor to the molecular pathogenesis of neurodegenerative diseases such as Parkinson's disease¹. Recent studies of the Parkinson's disease associated genes *PINK1* (ref. 2) and *parkin* (*PARK2*, ref. 3) indicate that they may act in a quality control pathway preventing the accumulation of dysfunctional mitochondria^{4–8}. Here we elucidate regulators that have an impact on parkin translocation to damaged mitochondria with genome-wide small interfering RNA (siRNA) screens coupled to high-content microscopy. Screening yielded gene candidates involved in diverse cellular processes that were subsequently validated in low-throughput assays. This led to characterization of TOMM7 as essential for stabilizing PINK1 on the outer mitochondrial membrane following mitochondrial damage. We also discovered that HSPAIL (HSP70 family member) and BAG4 have mutually opposing roles in the regulation of parkin translocation. The screens revealed that SIAH3, found to localize to mitochondria, inhibits PINK1 accumulation after mitochondrial insult, reducing parkin translocation. Overall, our screens provide a rich resource to understand mitochondrial quality control.

Following the loss of mitochondrial membrane potential, PINK1 and parkin coordinate a ubiquitination⁹, proteasomal activation¹⁰ and autophagic (mitophagy)⁵ response that may attenuate cell death¹¹. As mitophagy can ameliorate the deleterious consequences of mitochondrial dysfunction^{12–14}, genes regulating parkin translocation could be useful drug targets for increasing mitochondrial quality control. To identify genes important for the PINK1-dependent recruitment of parkin to damaged mitochondria, we conducted genome-wide siRNA screens against two diverse libraries. Our approach used the well-characterized cellular phenotype of parkin accumulation on depolarized mitochondria⁵. HeLa cells stably expressing GFP-parkin (Extended Data Fig. 1a) and a mitochondrial-targeted red fluorescent protein (mito-dsRed) were transfected with siRNA duplexes in 384-well plates (Extended Data Fig. 1b). After siRNA treatment, mitochondria were chemically depolarized with carbonyl cyanide *m*-chlorophenyl hydrazine (CCCP) to induce parkin translocation. Chemical depletion of mitochondrial membrane potential ($\Delta\psi$) mimicked pathological conditions caused by genetic mutation in mitochondrial- or nuclear-encoded genes or other stresses that deplete $\Delta\psi$. The degree of Parkin translocation was then assessed by high-content microscopy and automated image analysis that extracted a multitude of phenotypic parameters (Extended Data Figs 1c, d and 2a, b). *PINK1* siRNA-treated wells served as positive controls, abolishing parkin translocation following CCCP treatment (Fig. 1a). We observed a high degree of assay robustness ($Z' > 0.5$, Extended Data Fig. 2c, d) and technical reproducibility (Extended Data Fig. 2e, f) in primary screens. Assay termination before saturation (Extended Data Fig. 2g) allowed us to detect both translocation inhibitors and accelerators (Extended Data Fig. 2h). For example, *LMAN1* siRNAs accelerated parkin translocation, suggesting the gene negatively regulates this process (Fig. 1a). Our imaging-based assay was

used to screen genome-wide arrays of Ambion single (3 unique siRNAs per gene) and Dharmacon pooled (pool of 4 unique siRNAs per gene) reagents to maximize discovery potential.

Candidate gene selection used the robust statistical measure of median absolute deviation (MAD) to standardize siRNA activities from the screens (Fig. 1b, c). Normalized data were analysed using both parkin translocation and cytotoxicity (cell count) measures (see Supplementary Methods and Extended Data Fig. 3). Mito-dsRed intensity was used to identify siRNAs that depleted mitochondria (Extended Data Fig. 3). As important modulators of parkin translocation may have been masked by low-potency gene knockdowns, selection of active reagents at lower thresholds was feasible with our non-pooled siRNA data set, where we observed coincident activity of unique siRNAs with the same gene target. After filtering for cytotoxic or mitochondria depleting siRNAs, a gene was chosen as a 'candidate' modulator if at least two siRNAs from the non-pooled data set were active based on phenotypic evaluation of parkin translocation (generally >1.5 MAD or <-1.5 MAD, Extended Data Fig. 4a, b). Pooled siRNA screen candidate selection criteria were set to yield similar numbers of genes as the non-pooled siRNA screen ($\sim \pm 2$ MAD, see Supplementary Methods). The candidate list (Supplementary Table 1) had a gene selection rate (see Supplementary Methods) similar to other large-scale RNAi (RNA interference) screens¹⁵. Twenty-four candidates overlapped between the two screens (Fig. 1c, yellow dots) including an ubiquitin-conjugating enzyme (*UBE2J2*) and a member of the Hedgehog pathway (*HHAT*). Additionally, knockdown of the predicted LOC401052 strongly accelerated parkin translocation with all four reagents tested. The limited number of shared candidates was consistent with previous studies^{16,17}.

A large number of initial candidate selections were involved in gene expression (17% of non-pooled, 24% of pooled candidates). As PINK1 is very labile and essential for parkin translocation⁶, the high rate of activity from housekeeping gene siRNAs suggests they may influence *PINK1* expression. *PINK1* also may be a source of off-target effects that confound RNAi screens^{18–20}. The non-pooled siRNA screening data enabled us to systematically profile the miRNA-like effects of the siRNA seed sequences. Complementarity between the 5' end or 'seed' region of the siRNA guide strand and the 3' untranslated region (UTR) of unintended messenger RNAs is a driver of off-target behaviour^{20,21}. Seed sequences matching the 3' UTR of *PINK1* exerted a highly biased inhibitory effect ($P < 2.2 \times 10^{-16}$) on parkin translocation compared to all other siRNA seeds in the library (Extended Data Fig. 4c). We also observed that strongly inhibitory siRNAs (>2 MAD) had $\sim 10\%$ more matches to the 3' UTR of *PINK1* than the corresponding accelerator siRNAs (Extended Data Fig. 4d). In total, 10,935 siRNAs ($\sim 17\%$) in the non-pooled screen had at least one hexamer²² seed match to the *PINK1* 3' UTR. To combat off-target effects we used common seed analysis (CSA)²³. CSA exposes the activity of siRNA seed-based off-target effects by weighting each reagent against the population of siRNAs in the screen sharing the same seed (Extended Data Fig. 4e–g). The

¹Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA. ²Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, USA. ³NIH Center for Regenerative Medicine, Bethesda, Maryland 20892, USA.

*These authors contributed equally to this work.

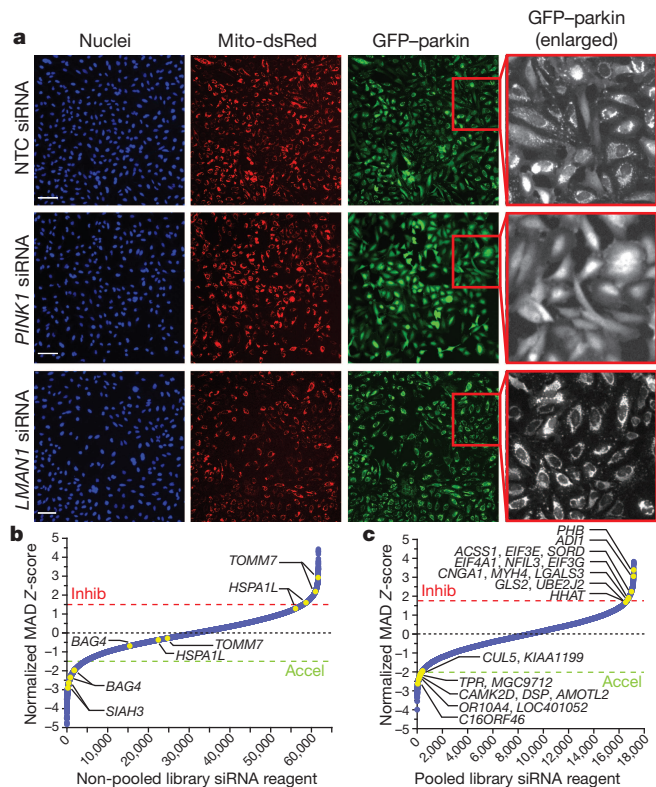


Figure 1 | A high-content screening assay for regulators of parkin translocation in HeLa cells following mitochondrial damage.

a, Representative images of parkin translocation assay. NTC (non-targeting control), *PINK1* (inhibited translocation) and *LMAN1* (accelerated translocation) siRNA transfected cells from genome-wide screens. Red-boxed panels are magnified views of GFP-parkin. Scale bars, 100 μ m. **b**, Negative control normalized, log transformed MAD Z-scores of all non-pooled reagents; data ordered from most negative to positive. Dashed red (inhibitors, Inhib) or green lines (accelerators, Accel) are screen-specific cutoffs (see Supplementary Methods) for active siRNA reagents. Yellow dots represent genes explored in this manuscript. **c**, Same as **b** but for the pooled RNAi screen; yellow dots represent overlap of candidate gene selections between pooled and non-pooled RNAi screens.

gene-level scoring of the entire non-pooled data set was adjusted for seed bias (Supplementary Table 2). Analysis of C911 mismatch controls²⁴ supported the hypothesis that siRNAs with strong seed bias (low seed-adjusted Z-score) modulate parkin translocation predominantly through seed-driven off-target activity (Extended Data Fig. 4h). In addition to CSA, we used gene pathway enrichment on the original non-pooled screen candidates (Supplementary Table 1) as a complementary method to identify promising genes. While confirming the strong influence of gene expression modifiers in our non-pooled screen candidates, enrichment analysis (Fig. 2a) revealed the significant ($P < 0.05$) presence of pathways including muscle function, the ubiquitin-proteasome, and autophagy (Supplementary Table 3). Furthermore, queries of the original non-pooled candidates (Supplementary Table 1) against annotated databases (gene ontology (GO) and human MitoCarta) revealed ubiquitin and mitochondrial processes (Supplementary Tables 4 and 5). Within enrichment groups, STRING database analysis (see Supplementary Methods) indicated many putative gene interactions that may regulate parkin (Fig. 2b, c).

To query candidate genes for follow-up, we chose from four distinct categories (Supplementary Table 6). These categories were top performing genes in the seed-adjusted primary screening data (category 1); candidates within a pathway enrichment group (category 2); top candidates from the pooled screen (category 3), and rational selection of candidates guided by gene annotations (category 4). In total, 106 positive regulators of parkin translocation were selected for validation

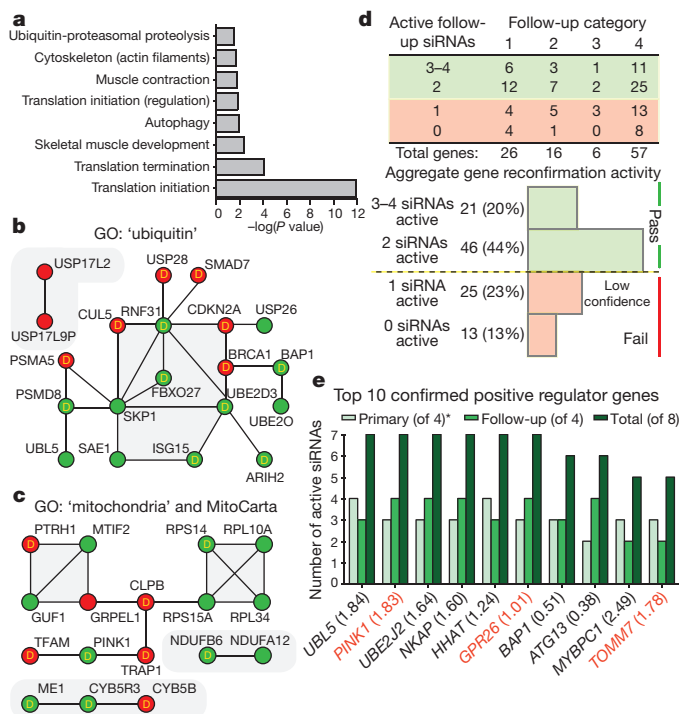


Figure 2 | Analysis of active siRNAs revealed linked gene clusters and sources of off-target effects. **a**, GeneGo cellular processes of candidate genes with significant enrichments (P -value cutoff of 0.05) (Supplementary Table 1). **b**, Candidate genes with ontology (GO) that included 'ubiquitin' and interactions by STRING association. Gene nodes are coloured green (positive regulator), red (negative regulator) and contain 'D' if they are members of the 'druggable' genome. **c**, Same as **b** but for the GO term 'mitochondria' or matches to the human MitoCarta database. **d**, Analysis of group specific (top panel) and aggregate (bottom panel) reconfirmation results (percentages are out of 105 tested). **e**, Top 10 positive regulator genes that passed reconfirmation (≥ 2 active siRNAs) in order of total active siRNAs and ties settled by seed-adjusted Z-score. Numbers in brackets indicate seed-adjusted Z-scores and red lettering indicates genes whose knockdown blocked PINK1 accumulation (see Supplementary Table 7). Asterisk indicates inclusion of siRNAs passing 'active' thresholds only in raw and non-transformed data sets because they exhibited excellent phenotypes upon inspection. Analyses in **a**–**c** were performed on original non-pooled candidate list (Supplementary Table 1) and **d**, **e** were performed on the candidates chosen for reconfirmation assays (Supplementary Table 6).

(Supplementary Table 6). Four additional siRNAs from a different vendor were used for high-throughput validation. We found 67 genes in follow-up studies recapitulated parkin translocation inhibition activity with at least two additional siRNA reagents (Fig. 2d), a common reconfirmation benchmark. This cutoff was further substantiated by quantitative PCR with reverse transcription (qRT-PCR) of randomly selected candidates with only 1 active follow-up siRNA (of 4). In all cases, equivalent knockdown of target mRNA was achieved by both parkin translocation inhibitory and non-active siRNAs in the set (Extended Data Fig. 5a–f). Therefore, on-target knockdown was unlikely to be source of the translocation phenotype in these low confidence reconfirmations. The 67 confirmed candidates were ranked by the number of active siRNA reagents out of the 8 assayed (4 from primary screens and 4 from follow-up analysis) and then by seed-adjusted Z-score (Top 10 in Fig. 2e and complete list in Supplementary Table 7). For 8 of the top-ranked genes, we examined mRNA levels in cells transfected with primary screen active siRNAs and observed $>75\%$ target knockdown in all but 1 of the 18 siRNAs tested (Extended Data Fig. 5g–n). Secondary screening of active follow-up siRNAs for PINK1 (expressed without its endogenous 3' UTR) immunofluorescence after CCCP treatment established if these gene knockdowns affected PINK1 protein accumulation (Fig. 2e, red, and Supplementary Tables 7 and 8).

TOMM7, a component of the protein translocase of the outer mitochondrial membrane (TOM) complex, was a reconfirmed candidate from the non-pooled screen and the sixty-third most potent positive regulator of parkin translocation in the seed-adjusted data set. Knockdown of this MitoCarta member in the pooled siRNA screen also showed a translocation deficit. Interestingly, a report on mammalian TOMM7 demonstrated that gene knockdown did not change the efficiency of mitochondrial protein import²⁵. Our *TOMM7* gene knockdowns in HeLa cells resulted in a reduction of GFP-parkin translocation ($P < 0.001$, Extended Data Fig. 6a, b) and a reduction of *TOMM7* mRNA levels ($P < 0.001$, Extended Data Fig. 6c). To confirm *TOMM7* siRNA activity was not an off-target effect, we generated a knockout HCT116 cell line using transcription activator-like effector nuclease (TALEN) mediated genome editing (see Supplementary Methods). Wild-type mRNA and protein was undetectable in the *TOMM7* knockout cell line (Extended Data Fig. 6d, e), but *PINK1* mRNA expression was unchanged (Extended Data Fig. 6f). Notably, the knockout of *TOMM7* abolished the translocation of YFP-parkin after CCCP treatment (Fig. 3a, b and Extended Data Fig. 6g). Expression of HA-tagged *TOMM7* in knockout cells restored the YFP-parkin translocation to near wild-type levels ($P < 0.001$ for partial and complete translocation). We examined if *TOMM7* knockout affects mitophagy downstream of parkin translocation. After 24 h exposure to CCCP, <5% of *TOMM7* knockout HCT116 cells underwent detectable mitophagy compared to ~90% of the wild-type cells ($P < 0.001$) (Fig. 3c and Extended Data Fig. 7a). We also observed a deficit in parkin-dependent degradation of MFN1 in knockout cells (Fig. 3d).

To explore TOMM7 function, we evaluated PINK1 levels in *TOMM7* wild-type and knockout HCT116 cell lines. PINK1 undergoes rapid turnover in polarized mitochondria⁶ and only trace amounts of PINK1 were detected in cell lysates from both lines (Fig. 3d). However, after CCCP treatment, normal accumulation of full-length PINK1 (ref. 6) failed to occur in *TOMM7* knockout cells, suggesting that TOMM7 is important for PINK1 stabilization on the outer mitochondrial membrane of damaged mitochondria. To explore why, we performed *in vitro* import of radiolabelled PINK1 protein into isolated mitochondria from wild-type or *TOMM7* knockout cells. Mitochondria from both cell lines were similar in their ability to import and process PINK1 precursor protein, as indicated by the amount of PARL-cleaved PINK1 (ref. 26) at each time point (Extended Data Fig. 7b, c). *In vitro* protein import and processing of the canonical Su9-DHFR precursor was also normal in *TOMM7* knockout mitochondria (Extended Data Fig. 7d, e). When $\Delta\psi$ is depleted, PINK1 associates with the TOM complex in the outer mitochondrial membrane²⁷. However, PINK1 imported *in vitro* into *TOMM7* knockout mitochondria failed to associate with the TOM complex (Fig. 3e, $-\Delta\psi$ lanes). Our results indicate that human TOMM7 functions in the TOM complex, not for generalized protein import, but to shunt and retain PINK1, and perhaps other proteins, to the surface of damaged mitochondria (Extended Data Fig. 7f).

To explore the wider importance of TOMM7, we depleted *TOMM7* gene expression in human induced pluripotent stem cell (iPS cell) derived neurons (Extended Data Fig. 7g–i). Lentiviral short hairpin RNA (shRNA) achieved a modest knockdown of *TOMM7* mRNA, without affecting *PINK1* mRNA levels (55–75%, Extended Data Fig. 7m, n). As in other cell types, significantly less endogenous PINK1 accumulated following mitochondrial depolarization in *TOMM7* knockdown neurons (Fig. 3f, g, $P < 0.001$ as compared to control). This suggests that TOMM7 functions to recruit parkin to damaged mitochondria by stabilizing PINK1 in neurons expressing tyrosine hydroxylase (TH⁺).

In search of positive regulators of parkin function, we were intrigued by *HSPA1L*, a confirmed follow-up selection (category 4, annotation/phenotype-based). As other HSP70 family members can interact with parkin^{28,29}, *HSPA1L* may selectively promote parkin translocation activity. *HSPA1L* is a widely distributed, but low-abundance member of the HSP70 family³⁰. The non-pooled siRNA screen yielded *BAG4* as a negative regulator of parkin translocation. BAG-domain-containing

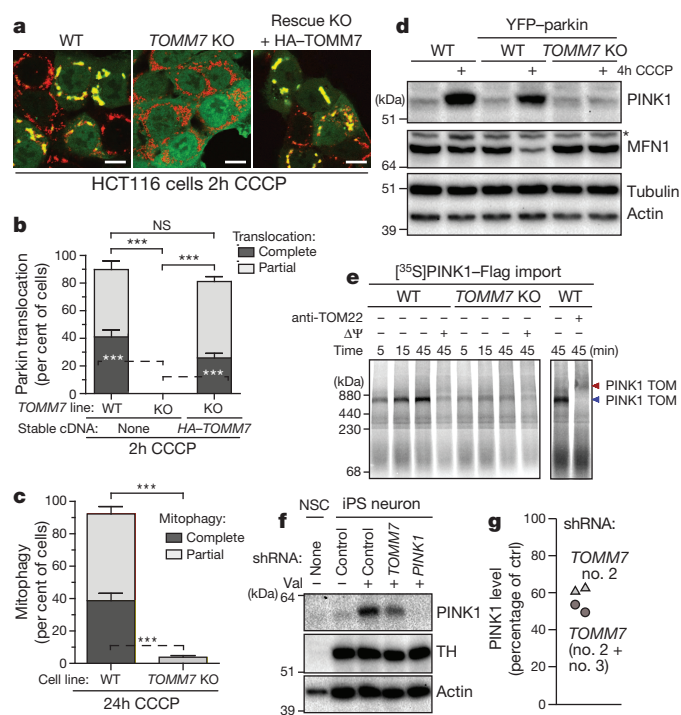


Figure 3 | Characterization of *TOMM7*, a positive regulator of parkin translocation.

a, Representative images of HCT116 cells stably expressing YFP-parkin (green) after 2 h of CCCP treatment: wild-type (WT), *TOMM7* knockout (KO), or *TOMM7* KO with stable expression of HA-TOMM7. Mitochondria are indicated by TOM20 immunofluorescence (red). **b**, Quantification of parkin translocation in **a**. **c**, Quantification of mitophagy observed using cells in **a** (24 h of CCCP) as measured by the loss of TOM20 and ATP5A signal (See Extended Data Fig. 7a). **d**, Western blot of whole-cell lysates from cells in **a** or wild-type HCT116 (lacking parkin) treated with CCCP for 0 or 4 h. An asterisk indicates a non-specific band. **e**, Blue native polyacrylamide gel electrophoresis (PAGE) import reactions (left) and super-shift (right) assay illustrate PINK1 association with the TOM complex. Blue arrow indicates PINK1 association with the TOM complex, red arrow indicates complex shifted by TOM22 antibody. **f**, Western blot of PINK1 and tyrosine hydroxylase (TH) levels in neuronal stem cells (NSC) and iPS-derived neurons. **g**, Quantification of PINK1 levels in **f**, expressed as the percentage of control levels ($n = 4$, $P < 0.001$ compared to control shRNA). Cells were infected with *TOMM7* shRNA no. 2 (triangles) or both no. 2 and no. 3 (circles). Quantification in **b** and **c** represent three independent experiments (>150 cells were counted per condition), are displayed as mean \pm s.d. and using one-way ANOVA tests (** $P < 0.001$). CCCP, 10 μM. Scale bar; 10 μm. For single channel images of **a** see Extended Data Fig. 6g.

proteins act as nucleotide exchange factors for HSP70 members and BAG5 can modulate the E3 ubiquitin ligase activity of parkin²⁹. We proposed that HSPA1L and BAG4 co-regulate parkin localization following mitochondrial damage. *HSPA1L* knockdown in HeLa cells led to a significant ($P < 0.01$) decrease in parkin translocation (Extended Data Fig. 8a–c) and *BAG4* knockdown enhanced parkin translocation ($P < 0.001$, Extended Data Fig. 8d–f). Neither of these knockdowns affected the level of PINK1 protein accumulation (Extended Data Fig. 8f, g). To ensure the phenotype was specific to *HSPA1L*, we used TALEN genome editing to knockout *HSPA1L* in HEK293 cells (see Supplementary Methods and Extended Data Fig. 8h). Parkin translocation was strongly inhibited in the *HSPA1L* knockout cells ($P < 0.001$, Fig. 4a, b) that have normal levels of HSPA1A (a homologous, more abundant HSP70), and equivalent levels of PINK1 (Extended Data Fig. 8i, j). This inhibition of parkin translocation was rescued by exogenous mCherry-*HSPA1L* (Fig. 4a, b and Extended Data Fig. 8k). Expression of exogenous *HSPA1A* did not rescue the *HSPA1L* knockout phenotype, indicating the translocation deficit is not from a loss of cytosolic chaperone capacity (Fig. 4b). To investigate the mechanism of *HSPA1L*

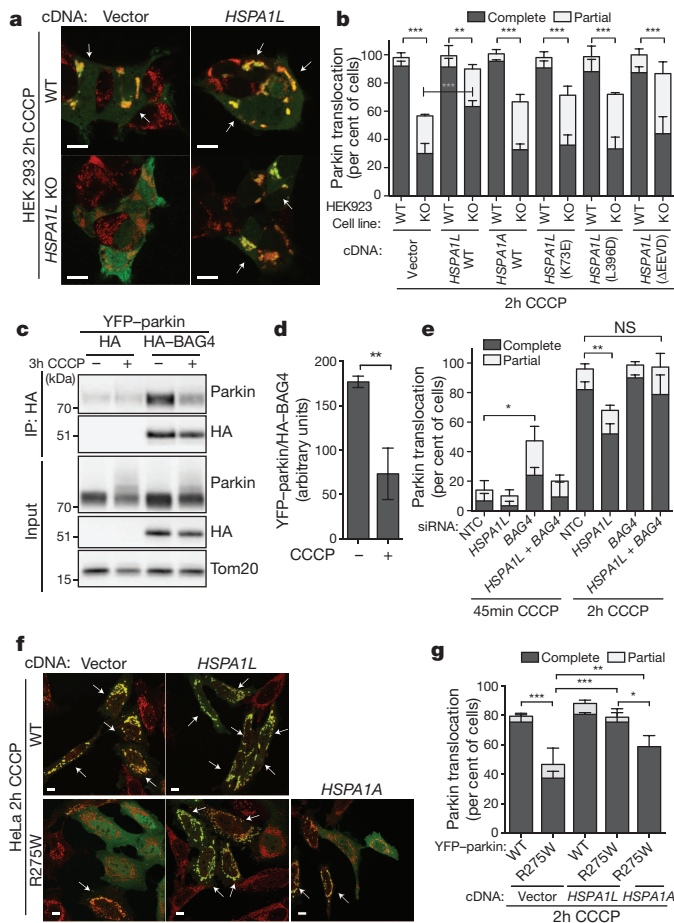


Figure 4 | HSPA1L and BAG4 differentially modulate translocation of parkin to damaged mitochondria. **a**, Representative images of wild-type (WT) and HSPA1L knockout (KO) HEK293 cells after 2 h CCCP with either mCherry vector or mCherry-HSPA1L overexpression showing rescue (YFP-parkin (green), Tom20 (red)). **b**, Quantification of parkin translocation from **a** and other rescue attempts. **c**, YFP-parkin is present in immunoprecipitates of HA-BAG4. **d**, Quantification of the intensity of bands in **c** ($n = 3$). **e**, siRNA knockdown of both *BAG4* and *HSPA1L* abrogates the phenotype of either individual gene knockdown. **f**, Translocation of parkin patient mutation R275W. **g**, Quantification of **b**, **e** and **g** represent three independent experiments (>150 cells were counted per condition), are mean \pm s.d.; one-way ANOVA tests ($*P < 0.05$, $**P < 0.01$ and $***P < 0.001$). Statistical comparisons are displayed for only complete translocation phenotype. CCCP, 10 μ M. White arrows indicate cells with complete parkin translocation. Scale bars, 10 μ m. For single channel images of **a** and **f** see Extended Data Figs 8k and 9g.

modulation of parkin, mutations were introduced into conserved domains: the ATPase domain (Lys73Glu), the Hsp40 binding domain (Leu396Asp), and a truncation in the regulatory/binding region of the carboxyl terminus (Δ EEVD) (see Supplementary Methods). None of these mutants rescued the translocation deficit of the *HSPA1L* knockout cells ($P < 0.01$, Fig. 4b). GFP-parkin immunoprecipitation from the HeLa cell line used in the screens was found to bind endogenous HSPA1L and other HSP70 isoforms by mass spectrometric analysis (Extended Data Fig. 8l and Supplementary Table 9). Immunoprecipitation of HA-BAG4 demonstrates that YFP-parkin also binds to BAG4 and this binding is diminished following mitochondrial depolarization ($P < 0.01$, Fig. 4c, d). The reciprocal immunoprecipitation of YFP-parkin also shows HA-BAG4 binding to both full-length and to the Δ UBL (ubiquitin-like domain) form of parkin, with stronger binding to the Δ UBL form (Extended Data Fig. 9a). A different BAG domain containing protein, BAG5, had been previously shown to interact with parkin²⁹. We confirmed this binding, but observed a greater binding to

HA-BAG4 (Extended Data Fig. 9b). RNAi knockdown of *BAG4* or *HSPA1L* alone promotes or inhibits parkin translocation, respectively. However, simultaneous knockdown of both genes in either HeLa (Fig. 4e and Extended Data Fig. 9c) or BE(2)-M17 neuroblastoma cell lines (Extended Data Fig. 9d, e) reduces their respective individual phenotypes. Therefore, neither protein is uniquely required for parkin translocation, but rather an imbalance between them causes robust phenotypes. BAG4 and HSPA1L appear to act together to regulate parkin translocation (Extended Data Fig. 9f).

We assessed the effect of HSPA1L on the translocation of disease-associated parkin mutants⁶. In HeLa cells *HSPA1L* enhanced the translocation of YFP-parkin(R275W) ($P > 0.001$, Fig. 4f, g and Extended Data Fig. 9g) to a greater extent than *HSPA1A*. Future studies on the molecular details of these interactions may provide attractive targets to enhance selective removal of damaged mitochondria.

As the siRNA retesting of negative parkin translocation regulators did not consistently recapitulate the generally more subtle acceleration phenotype, we used the seed-adjusted data set (Supplementary Table 2) to guide the pursuit of these genes. As the tenth most potent seed-adjusted negative regulator, the common seed plot for *SIAH3* siRNAs showed a consistent departure from their seed 'peers' (Extended Data Fig. 4g). *SIAH3*-Myc (*SIAH*/SINA superfamily member, see Extended Data Fig. 10a) co-localized with mitochondria (Extended Data Fig. 10b, c). We did not detect a stable interaction of *SIAH3* and PINK1 in the presence of detergents required for immunoprecipitation (Extended Data Fig. 10d). *SIAH3* knockdown in HeLa cells accelerated parkin translocation with an increase in partially translocated parkin at 30 ($P < 0.001$) and 60 ($P < 0.05$) minutes post CCCP treatment (Extended Data Fig. 10e, f) and complete translocation at 60 min ($P < 0.001$). RNAi rescue by *SIAH3* cDNA expression (siRNA-resistant) significantly ($P < 0.01$) suppressed the complete parkin translocation phenotype (Extended Data Fig. 10e, f). We confirmed the *SIAH3*-mediated parkin translocation phenotype ($P > 0.01$) and $>90\%$ knockdown of *SIAH3* mRNA in BE(2)-M17 neuroblastoma cells stably expressing parkin and mito-GFP ($P > 0.001$, Extended Data Fig. 10g-i). *SIAH3* silencing increased PINK1 accumulation after CCCP treatment ($P > 0.001$, Extended Data Fig. 10j, k). *SIAH3* probably acts as a negative regulator of PINK1 stabilization because knockdown of *SIAH3* did not increase *PINK1* mRNA or protein before mitochondrial damage (Extended Data Fig. 10j-l).

In conclusion, we used a diverse siRNA screening strategy to shed light on the genes regulating PINK1-dependent parkin translocation to damaged mitochondria. Using several analysis methods in parallel allowed us to identify and overcome the significant susceptibility of parkin translocation to off-target/pleiotropic *PINK1* modulation and select strong candidates for further study. Our subsequent validation demonstrated unique molecular functions of HSPA1L, BAG4 and *SIAH3*, as well as defining a role for TOMM7 in outer mitochondrial membrane PINK1 stabilization. Our genome-wide analysis of *PINK1* and parkin regulators illuminates mechanisms of mitochondrial maintenance.

METHODS SUMMARY

siRNA screening. siRNAs were arrayed in 384-well optical plates (2 μ l per well, 0.8 pmol). For each well, 20 μ l of DMEM medium (Life Technologies) containing 4 μ l ml⁻¹ RNAiMAX (Life Technologies) was added. After 30 min incubation, 20 μ l of a 37,500 cells per ml suspension was added to each well for reverse transfection. Plates with cells were incubated at 37 °C/5% CO₂ for 48 h and the parkin translocation assay was initiated with 40 μ l of DMEM containing CCCP (Sigma) at a final concentration of 10 μ M. After 2.5 h of incubation with CCCP, cells in all wells were fixed, nuclear stained and analysed using an ImageExpress Micro (Molecular Devices). All screening was performed in environmentally controlled enclosures with automated liquid and plate handling. See Supplementary Methods for detailed methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 June 2012; accepted 11 October 2013.

Published online 24 November 2013.

- Schapira, A. H. & Tolosa, E. Molecular and clinical prodrome of Parkinson disease: implications for treatment. *Nature Rev. Neurol.* **6**, 309–317 (2010).
- Valente, E. M. *et al.* Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. *Science* **304**, 1158–1160 (2004).
- Kitada, T. *et al.* Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
- Vincow, E. S. *et al.* The PINK1–Parkin pathway promotes both mitophagy and selective respiratory chain turnover *in vivo*. *Proc. Natl Acad. Sci. USA* **110**, 6400–6405 (2013).
- Narendra, D., Tanaka, A., Suen, D. F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803 (2008).
- Narendra, D. P. *et al.* PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol.* **8**, e1000298 (2010).
- Vives-Bauza, C. *et al.* PINK1-dependent recruitment of Parkin to mitochondria in mitophagy. *Proc. Natl Acad. Sci. USA* **107**, 378–383 (2010).
- Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
- Sarraf, S. A. *et al.* Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. *Nature* **496**, 372–376 (2013).
- Chan, N. C. *et al.* Broad activation of the ubiquitin-proteasome system by Parkin is critical for mitophagy. *Hum. Mol. Genet.* **20**, 1726–1737 (2011).
- Gautier, C. A., Kitada, T. & Shen, J. Loss of PINK1 causes mitochondrial functional defects and increased sensitivity to oxidative stress. *Proc. Natl Acad. Sci. USA* **105**, 11364–11369 (2008).
- Pimenta de Castro, I. *et al.* Genetic analysis of mitochondrial protein misfolding in *Drosophila melanogaster*. *Cell Death Differ.* **19**, 1308–1316 (2012).
- Suen, D. F., Narendra, D. P., Tanaka, A., Manfredi, G. & Youle, R. J. Parkin overexpression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. *Proc. Natl Acad. Sci. USA* **107**, 11835–11840 (2010).
- Burman, J. L., Yu, S., Poole, A. C., Decal, R. B. & Pallanck, L. Analysis of neural subtypes reveals selective mitochondrial dysfunction in dopaminergic neurons from *parkin* mutants. *Proc. Natl Acad. Sci. USA* **109**, 10438–10443 (2012).
- Sigoiillot, F. D. & King, R. W. Vigilance and validation: keys to success in RNAi screening. *ACS Chem. Biol.* **6**, 47–60 (2011).
- Bushman, F. D. *et al.* Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* **5**, e1000437 (2009).
- Simpson, J. C. *et al.* Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. *Nature Cell Biol.* **14**, 764–774 (2012).
- Schultz, N. *et al.* Off-target effects dominate a large-scale RNAi screen for modulators of the TGF- β pathway and reveal microRNA regulation of TGFB2. *Silence* **2**, 3 (2011).
- Sudbery, I., Enright, A. J., Fraser, A. G. & Dunham, I. Systematic analysis of off-target effects in an RNAi screen reveals microRNAs affecting sensitivity to TRAIL-induced apoptosis. *BMC Genomics* **11**, 175 (2010).
- Jackson, A. L. *et al.* Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).
- Birmingham, A. *et al.* 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods* **3**, 199–204 (2006).
- Buehler, E. *et al.* siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci. Reports* **2**, 428 (2012).
- Marine, S., Bahl, A., Ferrer, M. & Buehler, E. Common seed analysis to identify off-target effects in siRNA screens. *J. Biomol. Screen.* **17**, 370–378 (2012).
- Buehler, E., Chen, Y. C. & Martin, S. C911: A bench-level control for sequence specific siRNA off-target effects. *PLoS ONE* **7**, e51942 (2012).
- Kato, H. & Mihara, K. Identification of Tom5 and Tom6 in the preprotein translocase complex of human mitochondrial outer membrane. *Biochem. Biophys. Res. Commun.* **369**, 958–963 (2008).
- Jin, S. M. *et al.* Mitochondrial membrane potential regulates PINK1 import and proteolytic destabilization by PARL. *J. Cell Biol.* **191**, 933–942 (2010).
- Lazarou, M., Jin, S. M., Kane, L. A. & Youle, R. J. Role of PINK1 binding to the TOM complex and alternate intracellular membranes in recruitment and activation of the E3 ligase Parkin. *Dev. Cell* **22**, 320–333 (2012).
- Moore, D. J., West, A. B., Dikeman, D. A., Dawson, V. L. & Dawson, T. M. Parkin mediates the degradation-independent ubiquitination of Hsp70. *J. Neurochem.* **105**, 1806–1819 (2008).
- Kalia, S. K. *et al.* BAG5 inhibits parkin and enhances dopaminergic neuron degeneration. *Neuron* **44**, 931–945 (2004).
- Hageman, J. & Kampinga, H. H. Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. *Cell Stress Chaperones* **14**, 1–21 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Tuzmen for qRT-PCR assistance; C. Klumpp for automation; D. Maric for FACS sorting; C. Smith for microscopy assistance; H. Jaffe for mass spectrometry analyses; R. Fields (NINDS) for lentivirus assistance; N. Malik for iPS-derived neurons; the NINDS DNA Sequencing Facility; K. Mihara for the TOMM7 antibody and A. Koretsky and N. Caplen for support. Research was supported by the Japan Society for Promotion of Science Postdoctoral Fellowship for Research Abroad (K.Y.), the NIGMS Research Associate Program, the Intramural Research Program of the NIH, NINDS and the Trans-NIH RNAi initiative.

Author Contributions S.A.H. and R.J.Y. conceived the project; S.A.H., L.A.K., K.Y., D.A.S., C.-H.H., T.H., C.W., S.E.M. and R.J.Y. designed experiments; S.A.H., L.A.K., K.Y., S.M.H.-A., C.-H.H., D.A.S., T.H. and C.W. performed experiments; S.A.H., E.B., R.G. and S.E.M. conducted bioinformatics; and S.A.H. performed the screens. S.A.H., L.A.K. and R.J.Y. wrote the manuscript.

Author Information All screening data has been deposited into PubChem under the assay IDs 651811 (non-pooled screen, with siRNA sequences) and 651810 (pooled screen). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.Y. (youler@ninds.nih.gov).

p53 status determines the role of autophagy in pancreatic tumour development

Mathias T. Rosenfeldt¹, Jim O'Prey¹, Jennifer P. Morton¹, Colin Nixon¹, Gillian MacKay¹, Agata Mrowinska¹, Amy Au¹, Taranjit Singh Rai², Liang Zheng¹, Rachel Ridgway¹, Peter D. Adams², Kurt I. Anderson¹, Eyal Gottlieb¹, Owen J. Sansom¹ & Kevin M. Ryan¹

Macroautophagy (hereafter referred to as autophagy) is a process in which organelles termed autophagosomes deliver cytoplasmic constituents to lysosomes for degradation¹. Autophagy has a major role in cellular homeostasis and has been implicated in various forms of human disease^{2–4}. The role of autophagy in cancer seems to be complex, with reports indicating both pro-tumorigenic and tumour-suppressive roles^{3,5–12}. Here we show, in a humanized genetically-modified mouse model of pancreatic ductal adenocarcinoma (PDAC), that autophagy's role in tumour development is intrinsically connected to the status of the tumour suppressor p53. Mice with pancreases containing an activated oncogenic allele of *Kras* (also called *Ki-Ras*)—the most common mutational event in PDAC¹³—develop a small number of pre-cancerous lesions that stochastically develop into PDAC over time. However, mice also lacking the essential autophagy genes *Atg5* or *Atg7* accumulate low-grade, pre-malignant pancreatic intraepithelial neoplasia lesions, but progression to high-grade pancreatic intraepithelial neoplasias and PDAC is blocked. In marked contrast, in mice containing oncogenic *Kras* and lacking p53, loss of autophagy no longer blocks tumour progression, but actually accelerates tumour onset, with metabolic analysis revealing enhanced glucose uptake and enrichment of anabolic pathways, which can fuel tumour growth. These findings provide considerable insight into the role of autophagy in cancer and have important implications for autophagy inhibition in cancer therapy. In this regard, we also show that treatment of mice with the autophagy inhibitor hydroxychloroquine, which is currently being used in several clinical trials¹⁴, significantly accelerates tumour formation in mice containing oncogenic *Kras* but lacking p53.

To analyse the role of autophagy in pancreatic tumour development we generated mice that contained Cre recombinase driven by the *Pdx* promoter (which recombines in exocrine and endocrine pancreatic tissue¹⁵) and two floxed alleles of *Atg7* (ref. 16). As seen on recombination of other alleles by *Pdx-Cre* (ref. 15), loss of *Atg7* in the pancreas is mosaic (Fig. 1a). Concordantly, LC3 puncta—a marker of autophagosomes^{17,18}—were detected in *Atg7*-positive regions, whereas *Atg7*-negative regions had strong diffuse LC3 staining with occasional LC3 aggregates indicative of the cytoplasmic form of LC3 (LC3-I) that does not integrate into autophagosomes (Fig. 1a)¹⁷. *Atg7*-negative regions also accumulated the autophagic substrate p62 (also called SQSTM1) (Fig. 1a)¹⁹.

Analysis of these mice at various times after birth revealed no signs of pancreatic tumour formation or pre-malignancy; different to previous observations in the liver¹¹. However, progressive cellular destruction was observed in endocrine and exocrine pancreatic tissue with acinar tissue exhibiting significant accumulation of p53 (Fig. 1b, c and Extended Data Fig. 1a, b). Furthermore, activated caspase-3, a critical cell death factor downstream of p53, was also elevated in these regions (Fig. 1b, c and Extended Data Fig. 1a, b)²⁰. By contrast, both p53 and cleaved caspase-3 were virtually undetectable in wild-type tissue at all times (Fig. 1b and Extended Data Fig. 1a, b). Notably, p53 and caspase-3

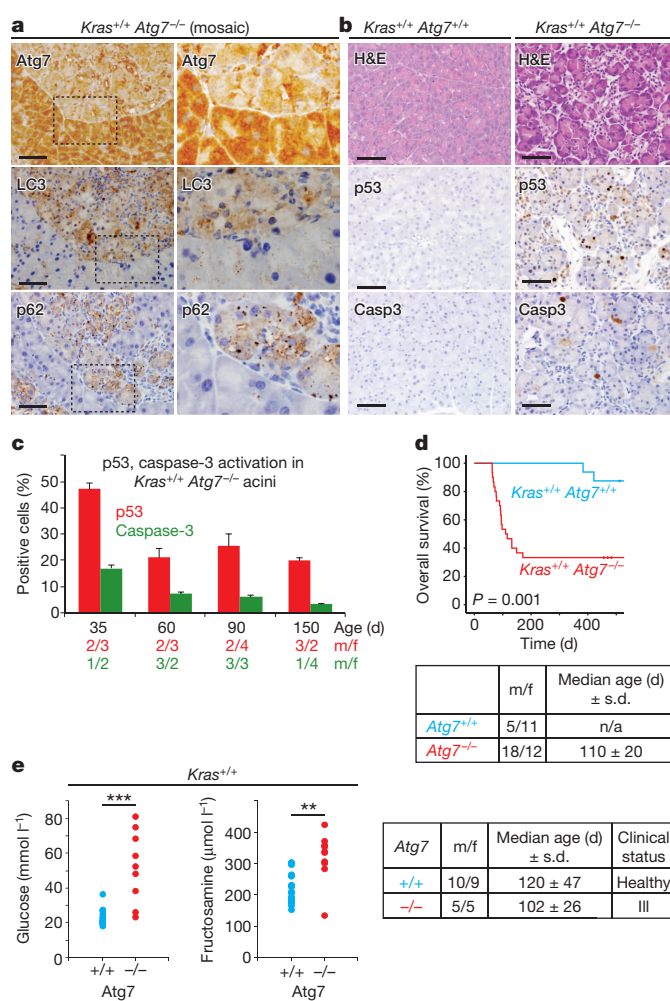


Figure 1 | Loss of autophagy in the pancreas causes organ destruction and reduces survival. **a**, Pancreatic sections stained with Atg7, LC3 and p62 showing *Atg7*^{-/-} next to *Atg7*^{+/+} regions after *Pdx-Cre*-mediated recombination. Right-side images are cropped and magnified regions (boxed) from left-side images. The magnification is $\times 2.5$. **b**, Haematoxylin and eosin (H&E), p53 and caspase-3 staining of exocrine tissue in *Atg7*^{+/+} and *Atg7*^{-/-} pancreases. **c**, Quantification (median, s.e.m.) of p53 and caspase-3 activation in *Atg7*-deficient acinar tissue. **d**, Kaplan–Meier analysis comparing overall survival of mice either wild type (blue) or *Atg7*^{-/-} (red) in the pancreas. Mantel–Cox test was used for statistics. ‘n/a’, 14/16 mice were killed >500 days when completely healthy. **e**, Biochemical analysis of endocrine function in moribund *Atg7*^{-/-} or age-matched healthy control mice. Detailed information about the mice is tabulated. A Mann–Whitney *U*-test was used for statistics. Scale bars: 50 μ m. ****P* < 0.01 and *****P* < 0.001.

¹Cancer Research UK Beatson Institute, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK. ²Institute of Cancer Studies, University of Glasgow, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK.

activation were negligible in *Atg7*^{-/-} islets despite substantial and progressive destruction of endocrine tissue (Extended Data Fig. 1b). This indicates that *Atg7* loss causes tissue destruction in both compartments, but p53 and caspase-3 activation seems tissue specific (Fig. 1b, c and Extended Data Fig. 1a, b).

Approximately 40% of *Atg7*-deficient mice had lifespans equivalent to wild-type mice. However, the remaining 60% became ill and had to be killed between 50 and 150 days (Fig. 1d). In line with endocrine tissue destruction, biochemical analyses of moribund *Atg7*^{-/-} mice revealed evidence of diabetes, but exocrine biochemistry was not affected despite destruction of this tissue being observed histologically (Fig. 1e and Extended Data Fig. 1a, c). This may be because exocrine parameters can be influenced by other tissues or that acinar cells, unlike β -cells, have considerable regenerative capacity²¹. Similar results were observed in mice lacking a second essential autophagy gene, *Atg5* (Extended Data Fig. 2).

p53 clearly has a role in these effects, as caspase-3 activation in acinar tissue is reduced and time of death, but not mortality rate, is significantly delayed by simultaneous deletion of p53 (Extended Data Fig. 3a–c). Notably, islet destruction still occurs in *Atg7*^{-/-} *p53*^{-/-} pancreases and these mice show signs of diabetes (Extended Data Fig. 3d). p53 deletion in the presence of *Atg7* over the same time frame had no effect (Extended Data Fig. 3e, f).

We next assessed the impact of *Atg7* deletion in mice expressing mutant *Kras* (with point mutation G12D) in their pancreases. In otherwise wild-type mice, *Kras*^{G12D} causes the appearance of a small number of pre-cancerous pancreatic intraepithelial neoplasia (PanIN), which stain with Alcian blue/PAS (Fig. 2a, b and Extended Data Fig. 4a, b)¹⁵. Progression to PDAC then occurs stochastically over time (Fig. 2c, d). In contrast, pancreases from mice with additional deletion of *Atg7* exhibited marked but transient accumulation of early PanINs (PanIN1A/B) (Fig. 2a, b, e and Extended Data Fig. 4a, c). In line with previous reports, this may be caused by secondary inflammation on tissue destruction,

which in our model is ultimately a consequence of *Atg7* deletion in pancreatic tissue^{12,20}. Notably, unlike PanINs in *Atg7*^{+/-} mice which progress to late-grade PanINs (PanIN2 and PanIN3), early PanINs in age-matched (150 days) *Atg7*-null mice undergo very limited progression to PanIN2 and virtually no progression to PanIN3, which precedes PDAC (Fig. 2e). Importantly, although mice containing mutant *Ras* and lacking *Atg7* ultimately became sick with signs of endocrine dysfunction, none showed histological signs of PDAC (Fig. 2c, d and Extended Data Figs 4c and 5a). This therefore shows that despite accumulation of early-grade PanINs, progression to late-grade PanIN and ultimately to PDAC does not occur during the lifespan of *Atg7*-deficient mice (Fig. 2d).

To understand why *Atg7*^{-/-} PanINs do not progress, we first determined if they arise from autophagy-deficient tissue. Lack of LC3 puncta and accumulation of p62 confirmed that this was the case (Extended Data Fig. 4a). PanINs also had elevated p53 levels irrespective of *Atg7* status, and those lacking *Atg7* had markedly enhanced caspase-3 activation (Fig. 2f and Extended Data Figs 4b, c and 5b, c). p53 activation in *Ras*-driven pancreatic lesions causes growth arrest and cellular senescence, which act as barriers to tumour development^{22,23}. *In vitro* studies have reported, however, that timely establishment of senescence requires autophagy²⁴, but autophagy-deficient PanINs from *Atg7*^{fl/fl} mice show sustained growth arrest as judged by Ki67 staining, and stain strongly for the senescence markers senescence-associated β -galactosidase and p21 (Extended Data Figs 4c and 5b, d, e). We conclude, therefore, that growth arrest, senescence and cell death act as barriers to the progression of PanINs to PDAC in the absence of *Atg7*.

Because *Atg7* has an autophagy-independent role and because early death limited analysis of PDAC progression in *Atg7*^{-/-} mice, we also assessed PDAC development in the absence *Atg5* (refs 25, 26). PanIN formation was again observed when mutant *Kras* was expressed in *Atg5*-deficient pancreases (Extended Data Fig. 6), but progression to PDAC never occurred even though some mice had lifespans over 500 days of

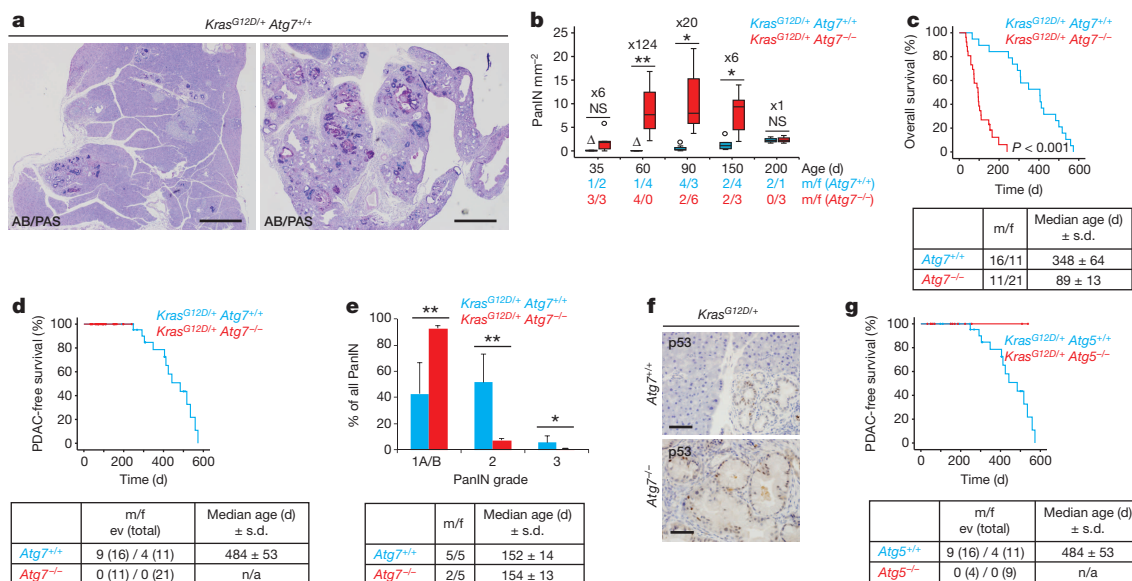


Figure 2 | Autophagy-deficient mice show enhanced Ras-driven PanIN formation but do not develop PDAC. **a**, Alcian blue/PAS (AB/PAS) staining in pancreases from 150-day-old *Atg7*-proficient or *Atg7*-deficient *Kras*^{G12D/+} mice (wild-type mice with expression of *Kras*^{G12D} in their pancreases). **b**, Quantification of PanIN from mice of the indicated genotype and age. Fold increase and male/female ratios are shown. Box plots express the first (Q1) and third (Q3) quartiles set by the upper and lower limits of the box, whereas the median is shown by the horizontal line inside the box. The whiskers extend upward and downward to the highest or lowest observation within the upper (Q3 + 1.5 × the interquartile range) and lower (Q1 - 1.5 × interquartile range) limits. Values outside the upper and lower

limits are 'outliers' and are shown by individual symbols. **c**, **d**, Kaplan-Meier analysis comparing overall survival (**c**) or PDAC-free survival (**d**) of *Atg7*^{+/+} (blue) or *Atg7*^{-/-} (red) *Kras*^{G12D/+} mice. **e**, PanIN grading of pancreases from 150-day-old *Atg7*^{+/+} or *Atg7*^{-/-} Pdx-Cre *Kras*^{G12D/+} mice. Values are mean and error bars are s.d. **b**, **e**, A Mann-Whitney *U*-test was used for statistics. **f**, Representative p53 staining from 150-day-old mice of both genotypes. **g**, PDAC-free survival of autophagy-proficient (blue) and *Atg5*-deficient (red) *Kras*^{G12D/+} mice. **d**, **g**, Median survival ± s.d. and male/female (m/f) ratio are provided. 'n/a', no animals succumbed to PDAC. A log-rank test (Mantel-Cox) was used for statistics. **a**, **f**, Scale bars: 500 μ m (**a**), 50 μ m (**f**). **P* < 0.05, ***P* < 0.01; NS, not significant.

age (Fig. 2g). This therefore confirms that mutant Ras-driven PDAC requires autophagy in the presence of wild-type p53.

Loss or mutation of p53 occurs in 50% of human PDAC¹⁵. We therefore questioned whether p53 is required to block PanIN to PDAC progression upon deletion of *Atg7*. Mice were generated that contained a mutant allele of *Kras*, homozygous floxed alleles for *p53* and either wild-type or floxed alleles for *Atg7*. Not only did tumours now form irrespective of autophagy status, but time of death was significantly accelerated in *Atg7*^{-/-} compared to *Atg7*^{+/+} mice (Fig. 3a). All tumours arising in *Atg7*-deficient pancreases lacked *Atg7* expression and had accumulation of LC3-I and p62, indicating loss of autophagy (Fig. 3b).

We next considered whether mice lacking *Atg7* die early from accelerated tumour onset or pancreatic dysfunction. Mice were therefore killed at set times after birth and the mass of evident tumours recorded. This revealed that *Atg7* loss accelerates PDAC formation and age-matched tumours had a greater mass in *Atg7*^{-/-} mice (Fig. 3c, d). By contrast, no mice showed signs of pancreatic insufficiency, indicating that this was not the cause of early death in the absence of *Atg7* (Extended Data Fig. 7a). In conclusion, whereas *Atg7* is required for tumour development in a p53-proficient background, in the absence of p53, *Atg7* conversely has a tumour-suppressive role. Similar results were observed on deletion of *Atg5* (Fig. 3e and Extended Data Fig. 7b).

Inhibition of autophagy with hydroxychloroquine is considered a therapeutic option for cancer, and its use against mouse PDACs that developed with wild-type p53 and pancreatic tumour xenografts containing mutant p53 is reported to be beneficial^{9,14}. Our genetic data suggest, however, that hydroxychloroquine might instead be tumour-promoting in Ras-driven pancreatic tumours that develop without p53. To test this we treated autophagy-competent mice with hydroxychloroquine and found that PDAC formation was significantly accelerated (Fig. 3f and Extended Data Fig. 8a). Notably, treatment did not affect

pancreatic function or PanIN numbers in *Kras*^{G12D/+} *p53*^{+/+} mice (Extended Data Fig. 8b–d). These data therefore support that loss of *Atg7* or *Atg5* affects tumour development through autophagy inhibition and highlight caution in the use hydroxychloroquine in certain tumour settings.

Because autophagy is critical for tumour development driven by Ras alone, why is this not the case in the absence of p53? We noted that tumours from p53-deficient mice had markedly fewer autophagosomes when compared to those that developed with p53 (Fig. 4a). Similarly, cell lines from *p53*^{-/-} tumours had significantly lower levels of LC3-II (the autophagosome-bound form of LC3) and decreased autophagic flux (Fig. 4b and Extended Data Fig. 9a, b), indicating that autophagy is not actively promoted in these tumours/cell lines because their viability is not intricately dependent on the process. In studies where growth of Ras-driven tumours is autophagy-dependent, loss of autophagy causes decreased oxygen consumption—an indicator of diminished oxidative metabolism^{8,9}. In the absence of p53, however, loss of autophagy does not reduce oxygen consumption or mitochondrially associated metabolites such as citrate (Fig. 4c and Extended Data Fig. 9c), indicating that the TCA cycle is unaffected by autophagy loss in this context. In contrast, loss of autophagy causes increased extracellular acidification via enhanced glycolysis, as exemplified by increased glucose consumption, extracellular lactate accumulation and the fact that increased acidification can be blocked with the hexokinase inhibitor 2-deoxyglucose (Fig. 4d–g)²⁷.

We were also interested to understand why *Atg7*-deficient tumours manifest earlier, as analysis of pancreases at 21 days of age revealed equal PanIN numbers irrespective of *Atg7* status (Extended Data Fig. 9d). Our *in vitro* analysis revealed that autophagy-deficient cells have enhanced glucose uptake as measured by intake of 2-deoxyglucose (Fig. 4h), and analysis of ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) uptake

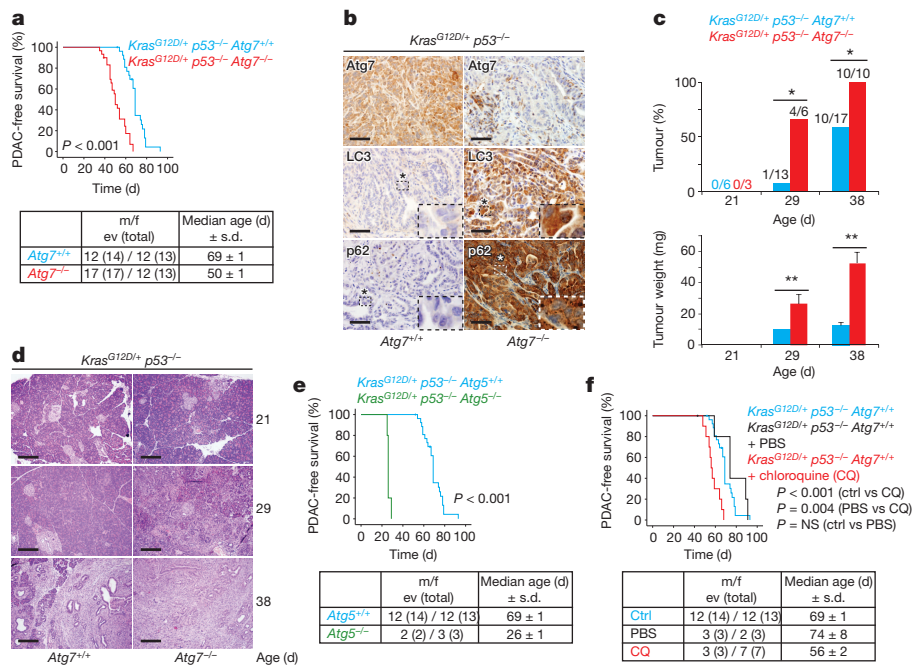


Figure 3 | Genetic or pharmacological inhibition of autophagy in the absence of p53 permits and accelerates PDAC formation. **a**, Kaplan-Meier analysis comparing PDAC-free survival in *Atg7*^{+/+} (blue) or *Atg7*^{-/-} (red) *Kras*^{G12D/+} *p53*^{-/-} mice. **b**, Pancreatic tumours of indicated genotypes stained for *Atg7*, LC3 and p62. Insets are magnified crops from the same images to show specifics of LC3 and p62 staining. The magnification is × 4. **c**, Tumour incidence and tumour weight (median, s.e.m.) in mice from indicated genotypes, killed at 21, 29 and 38 days of age. Numbers in the incidence diagram represent tumour-bearing mice versus all mice. Significance was assessed with Fisher's exact test (incidence) and Mann-Whitney U-test

(weight). **d**, Representative haematoxylin and eosin stainings of pancreases from indicated time points and genotypes. **e**, Kaplan-Meier analysis comparing PDAC-free survival of *Atg5*^{+/+} (blue) or *Atg5*^{-/-} (green) *Kras*^{G12D/+} *p53*^{-/-} mice. **f**, Kaplan-Meier analysis comparing PDAC-free survival of *Kras*^{G12D/+} *p53*^{-/-} mice that were chloroquine-treated (CQ, red), vehicle-treated (PBS, black) or untreated (blue). **a**, **e**, **f**. The blue curve in all Kaplan-Meier plots represents the same colony of mice (*Kras*^{G12D/+} *p53*^{-/-}, *Atg7*^{+/+}, *Atg5*^{+/+}). A Mantel-Cox test was used for statistics. **b**, **d**, Scale bars: 100 μm (**b**), 200 μm (**d**). **P* < 0.05, ***P* < 0.01.

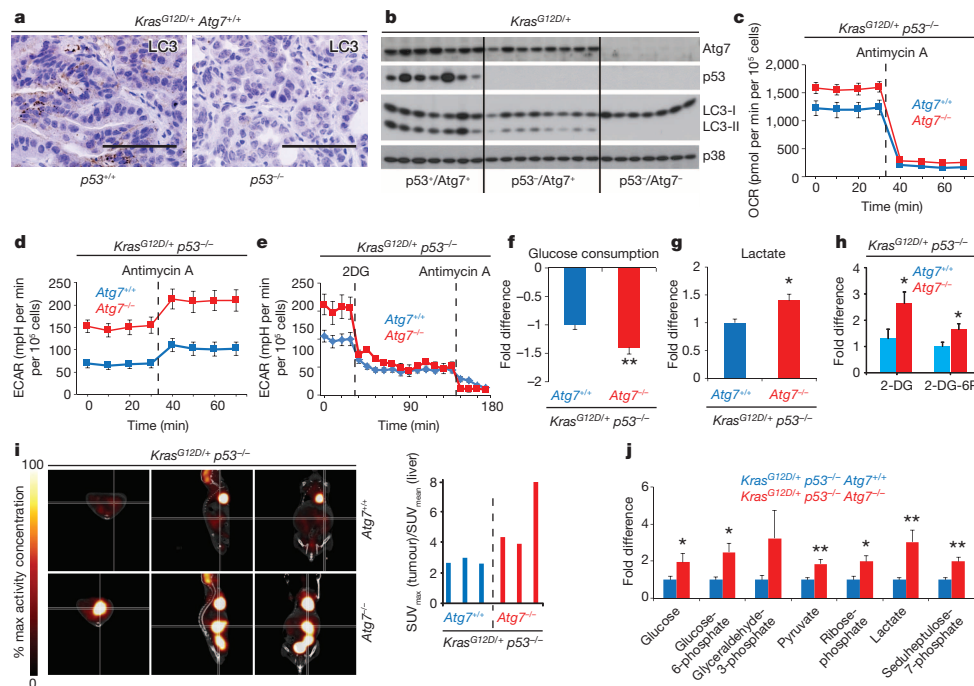


Figure 4 | Loss of p53 results in lower levels of autophagy, and combined loss of p53 and Atg7 enhances levels of glycolytic and pentose phosphate pathway intermediates. **a**, LC3 staining of representative pancreatic tumours from mice of indicated genotypes. Scale bars: 50 μ m. **b**, Western blotting for LC3, p53 and Atg7 in cell lines derived from individual tumours which developed in the absence or presence of p53. **c**, **d**, Average oxygen consumption rate (OCR, **c**) and extracellular acidification rate (ECAR, **d**) of eight *Atg7*^{+/+} and seven *Atg7*^{-/-} cell lines revealed increased ECAR in autophagy-deficient cells. Error bars are s.e.m. **e**, Treatment with 2-deoxyglucose (2-DG) reduces ECAR and abrogates the difference between cell lines from *Atg7*^{+/+} and *Atg7*^{-/-} tumours. Values are the average of three biological replicates. Error bars are s.e.m. **f**, **g**, LC-MS analysis revealed increased glucose consumption from medium (**f**) and extracellular lactate

accumulation (**g**) in *Atg7*-null cell lines. Values are the average of eight (*Atg7*^{+/+}) and seven (*Atg7*^{-/-}) biological replicates. **h**, LC-MS analysis shows increased accumulation of 2-DG and its metabolite 2-DG-6-phosphate in *Atg7*-deficient cell lines. Values are averages of five biological replicates for each genotype. **i**, Background levels of ¹⁸F-FDG uptake in brain and heart can be seen in sagittal and coronal views (middle and right panels, respectively) of wild-type *Atg7* mice (top row). In contrast, intense ¹⁸F-FDG uptake is seen in pancreatic tumours of *Atg7* deficient mice (bottom row). **j**, LC-MS of intracellular glycolytic and pentose phosphate pathway intermediates shows enhanced accumulation in *Atg7*^{-/-} tumour cell lines. Values are average of eight (*Atg7*^{+/+}) and six (*Atg7*^{-/-}) biological replicates. **f**–**h**, **j**, Error bars are s.e.m. **a**, **f**–**h**, **j**, A Mann–Whitney *U*-test was used to ascertain significance. **P* < 0.05, ***P* < 0.01.

by positron emission tomography confirmed that this was also the case *in vivo* (Fig. 4i).

Although autophagy loss causes DNA damage which could facilitate tumour growth^{28,29}, our data caused us to speculate that enhanced glucose uptake might also support anabolic pathways in autophagy-deficient cells. Analysis of endometabolites revealed that *Atg7*-null cells have higher levels of intracellular glucose, glycolytic intermediates and components of the pentose phosphate pathway (Fig. 4j). Owing to the role of the pentose phosphate pathway in the generation of reductive power and intermediates of amino and nucleic acid biosynthesis, we believe that autophagy loss in Ras-driven, p53-deficient tumours results in increased glucose uptake and enhanced pools of anabolic intermediates that are known to facilitate tumour growth²⁷. Interestingly, increased pentose phosphate pathway intermediates were not observed upon *Atg7* loss in cultures of pre-cancerous ductal tissue (Extended Data Fig. 9e–g). However, in PDACs these metabolic changes seem permanent as they could not be reversed on reconstitution of autophagy by ectopic expression of *Atg7* (Extended Data Fig. 10). This indicates that the changes in metabolism we observed in PDACs caused by deletion of *Atg7* are associated with p53 loss and may be a response to increased aggressiveness of the tumour. These findings clearly show that p53 status switches the role of autophagy during pancreatic tumour development, highlighting important considerations for the treatment of malignant disease.

METHODS SUMMARY

Mouse strains were described previously^{15,16,26} and all experiments were carried out in compliance with UK Home Office regulation. Where indicated, 3–4-week-old mice were treated with 60 mg kg⁻¹ day⁻¹ hydroxychloroquine by intraperitoneal

injection. For all Kaplan–Meier analyses shown in figures, tables are provided that show median survival \pm standard deviation (s.d.) and male/female ratios (m/f). In some cases, the number of mice succumbing to PDAC is shown (event (ev)), followed by the total number (total) of mice of the same genotype. For immunohistochemical analysis, samples were fixed for at least 24 h in neutral-buffered formalin and stained with the following antibodies as previously described¹⁸: p53 (Vector labs), p21 (Santa Cruz), LC3 (Nanotools), p62 (Biomol), ATG7 (Santa Cruz). Mucin from PanINs was visualized by staining with Alcian blue/Periodic acid¹⁵. For PanIN quantification all PanINs per histological section (H&E) were counted and divided by the area occupied by the pancreatic tissue on the slide. Cell lines from PDACs were generated and cultured as described previously²². Metabolic parameters were assessed from cells during logarithmic growth. Antimycin A was added where indicated to block mitochondrial respiration. Endo- and exometabolites were measured by LC-MS as previously described³⁰. Quantification of ¹⁸F-FDG uptake in *Atg7* wild-type and deficient mice is expressed as a ratio of the maxim standard uptake value (SUV) within the tumour to the mean SUV within a reference tissue (liver). Western blotting was undertaken using standard techniques. Proteins were detected with the following antibodies: p38 (Cell Signaling), LC3B (Cell Signaling), p53 (Cell Signaling), ATG7 (Santa Cruz). Where indicated, cells were treated with 10 mM 2DG. Cell counting and densitometry were done using ImageJ software. Tests used to determine statistical significance are provided in the figure legends. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 May 2012; accepted 13 November 2013.

Published online 4 December 2013.

1. Yang, Z. & Klionsky, D. J. Mammalian autophagy: core molecular machinery and signaling regulation. *Curr. Opin. Cell Biol.* **22**, 124–131 (2010).

2. Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451**, 1069–1075 (2008).
3. Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132**, 27–42 (2008).
4. Kundu, M. & Thompson, C. B. Autophagy: basic principles and relevance to disease. *Annu. Rev. Pathol.* **3**, 427–455 (2008).
5. Mah, L. Y. & Ryan, K. M. Autophagy and cancer. *Cold Spring Harb. Perspect. Biol.* **4**, a008821 (2012).
6. Yue, Z., Jin, S., Yang, C., Levine, A. J. & Heintz, N. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proc. Natl Acad. Sci. USA* **100**, 15077–15082 (2003).
7. Liang, X. H. *et al.* Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* **402**, 672–676 (1999).
8. Guo, J. Y. *et al.* Activated Ras requires autophagy to maintain oxidative metabolism and tumorigenesis. *Genes Dev.* **25**, 460–470 (2011).
9. Yang, S. *et al.* Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25**, 717–729 (2011).
10. Wei, H. *et al.* Suppression of autophagy by FIP200 deletion inhibits mammary tumorigenesis. *Genes Dev.* **25**, 1510–1527 (2011).
11. Takamura, A. *et al.* Autophagy-deficient mice develop multiple liver tumors. *Genes Dev.* **25**, 795–800 (2011).
12. Guo, J. Y. *et al.* Autophagy suppresses progression of K-ras-induced lung tumors to oncocytomas and maintains lipid homeostasis. *Genes Dev.* **27**, 1447–1461 (2013).
13. Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N. & Depinho, R. A. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev.* **20**, 1218–1249 (2006).
14. Amaravadi, R. K. *et al.* Principles and current strategies for targeting autophagy for cancer treatment. *Clin. Cancer Res.* **17**, 654–666 (2011).
15. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).
16. Komatsu, M. *et al.* Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *J. Cell Biol.* **169**, 425–434 (2005).
17. Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).
18. Rosenfeldt, M., Nixon, C., Liu, E., Mah, L. Y. & Ryan, K. M. Analysis of macroautophagy by immunohistochemistry. *Autophagy* **8**, 963–969 (2012).
19. Bjorkoy, G. *et al.* p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *J. Cell Biol.* **171**, 603–614 (2005).
20. Crichton, D. & Ryan, K. M. Splicing DNA-damage responses to tumour cell death. *Biochim. Biophys. Acta* **1705**, 3–15 (2004).
21. Desai, B. M. *et al.* Preexisting pancreatic acinar cells contribute to acinar cell, but not islet beta cell, regeneration. *J. Clin. Invest.* **117**, 971–977 (2007).
22. Morton, J. P. *et al.* Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc. Natl Acad. Sci. USA* **107**, 246–251 (2010).
23. Caldwell, M. E. *et al.* Cellular features of senescence during the evolution of human and murine ductal pancreatic cancer. *Oncogene* **31**, 1599–1608 (2012).
24. Young, A. R. *et al.* Autophagy mediates the mitotic senescence transition. *Genes Dev.* **23**, 798–803 (2009).
25. Lee, I. H. *et al.* Atg7 modulates p53 activity to regulate cell cycle and survival during metabolic stress. *Science* **336**, 225–228 (2012).
26. Hara, T. *et al.* Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. *Nature* **441**, 885–889 (2006).
27. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
28. Mathew, R. *et al.* Autophagy suppresses tumor progression by limiting chromosomal instability. *Genes Dev.* **21**, 1367–1381 (2007).
29. Karantza-Wadsworth, V. *et al.* Autophagy mitigates metabolic stress and genome damage in mammary tumorigenesis. *Genes Dev.* **21**, 1621–1635 (2007).
30. Frezza, C. *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225–228 (2011).

Acknowledgements We are grateful to M. Komatsu for providing the Atg7 floxed mice that were used in this study, and we thank M. O'Prey for help with fluorescent microscopy; G. Gillen for help with PET; T. McBryan for help with statistics; C. Frezza for help with metabolic studies; and A. Rosenwald for help and advice on pathology. This work was supported by Cancer Research UK and the Pancreatic Cancer Research Fund.

Author Contributions M.T.R. and K.M.R. conceived the study. M.T.R., E.G., P.D.A., K.I.A., O.J.S. and K.M.R. designed experiments. M.T.R., J.O.P., J.P.M., C.N., G.M., A.M., A.A., T.S.R., L.Z. and R.R. conducted experiments. M.T.R., J.P.M., A.R., E.G., P.D.A., K.I.A., O.J.S. and K.M.R. analysed experiments. M.T.R. and K.M.R. wrote the manuscript. J.O.P. and J.P.M. contributed equally to this study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.R. (k.ryan@beatson.gla.ac.uk).

Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1

Ioannis Manolaridis^{1*}, Kiran Kulkarni^{1*†}, Roger B. Dodd^{1†}, Satoshi Ogasawara^{2,3}, Ziguu Zhang¹, Ganka Bineva⁴, Nicola O'Reilly⁴, Sarah J. Hanrahan¹, Andrew J. Thompson¹, Nora Cronin¹, So Iwata^{2,3,5} & David Barford¹

CAAX proteins have essential roles in multiple signalling pathways, controlling processes such as proliferation, differentiation and carcinogenesis¹. The ~120 mammalian CAAX proteins function at cellular membranes and include the Ras superfamily of small GTPases, nuclear lamins, the γ -subunit of heterotrimeric GTPases, and several protein kinases and phosphatases². The proper localization of CAAX proteins to cell membranes is orchestrated by a series of post-translational modifications of the carboxy-terminal CAAX motifs³ (where C is cysteine, A is an aliphatic amino acid and X is any amino acid). These reactions involve prenylation of the cysteine residue, cleavage at the AAX tripeptide and methylation of the carboxyl-prenylated cysteine residue. The major CAAX protease activity is mediated by Rce1 (Ras and a-factor converting enzyme 1), an intramembrane protease (IMP) of the endoplasmic reticulum^{4,5}. Information on the architecture and proteolytic mechanism of Rce1 has been lacking. Here we report the crystal structure of a *Methanococcus maripaludis* homologue of Rce1, whose endopeptidase specificity for farnesylated peptides mimics that of eukaryotic Rce1. Its structure, comprising eight transmembrane α -helices, and catalytic site are distinct from those of other IMPs. The catalytic residues are located ~10 Å into the membrane and are exposed to the cytoplasm and membrane through a conical cavity that accommodates the prenylated CAAX substrate. We propose that the farnesyl lipid binds to a site at the opening of two transmembrane α -helices, which results in the scissile bond being positioned adjacent to a glutamate-activated nucleophilic water molecule. This study suggests that Rce1 is the founding member of a novel IMP family, the glutamate IMPs.

Rce1 is a type II CAAX prenyl endopeptidase that was first identified in *Saccharomyces cerevisiae* together with the type I CAAX-processing enzyme, Ste24p (also known as ZMPSTE24 or Afc1p)^{4,6}. This processing enzyme is a zinc metalloproteinase with a specific role in the processing of prelamin A in all eukaryotes and a-factor in yeast. Rce1, by contrast, has a much wider specificity, processing all farnesylated and geranylgeranylated CAAX proteins. However, extensive sequence and biochemical analyses have been unable to classify Rce1 within the three conventional IMP families. These are the rhomboids⁷, the intramembrane metalloproteinases (which include S2P)⁸ and the aspartyl proteases (which include presenilin⁹ and SPP). The membrane proteases ZMPSTE24 (refs 10, 11) and FlaK¹² have their catalytic sites at the membrane interface. Rce1 belongs to the ABI (abortive infection) family of putative IMPs and has homologues in all three domains of life. The ABI family is defined by three conserved motifs^{13,14} that constitute the catalytic site of the ABI proteases and whose importance has been demonstrated by mutational analysis of yeast Rce1p^{15,16}. Rce1 inactivation has also been shown in mice to result in the mislocalization of Ras proteins from the plasma membrane¹⁷. The consequent disruption of Ras signalling inhibited the Ras-induced transformation of fibroblasts¹⁷ but accelerated progression of K-Ras-induced myeloproliferative disease^{17,18}.

In addition, Rce1-deficient mice have been found to develop lethal dilated cardiomyopathy¹⁹, and Rce1 is also essential for the survival of photoreceptor cells in mice²⁰.

To understand the structure and catalytic mechanism of Rce1, we examined the expression and solubility properties of ~30 Rce1 homologues (including human, yeast and prokaryotic homologues) using fluorescence-detection size-exclusion chromatography (FSEC) and differential scanning fluorimetry (data not shown). Rce1 from the archaea *Methanococcus maripaludis* (MmRce1) was identified as a suitable candidate for structural studies, and the full-length protein (276 residues, 15% sequence identity with human RCE1) was crystallized in complex with a conformation-sensitive monoclonal antibody Fab fragment (Extended Data Fig. 1). The structure of MmRce1–Fab was determined by molecular replacement using the Fab fragment as a search model, and the complex was refined to 2.5 Å resolution (Extended Data Table 1).

A fluorescence-based protease assay showed that MmRce1 hydrolyses a farnesylated peptide modelled on the C terminus of human RhoA (Fig. 1a, b and Extended Data Fig. 2). Similar to its eukaryotic orthologues^{21,22}, peptide hydrolysis depends on a farnesylated cysteine residue; however, in contrast to human RCE1 (ref. 21), MmRce1 did not proteolytically cleave geranylgeranylated peptides (data not shown). Eukaryotic Rce1 is an endoprotease that specifically cleaves immediately C-terminally to the farnesyl cysteine (at position 1 (P1))^{4,23}. Mass spectrometry analysis indicated that MmRce1 is also an endoprotease, although with a slightly relaxed specificity, cleaving the CAAX motif C-terminal to both P1 and P1' (Fig. 1b, c). MmRce1 is inhibited by N-acetyl-S-farnesyl-L-cysteine (AFC), the minimal analogue of farnesylated peptides (Fig. 1a). The ability of MmRce1 to cleave specifically farnesylated peptides validates it as a model for understanding the mechanism of CAAX processing by eukaryotic Rce1.

The structure of MmRce1 comprises eight conserved transmembrane α -helices (TM1–8), with two peripheral membrane α -helices (α A and α B) (Fig. 2a, b). Rce1 is topologically distinct from other IMPs and, to our knowledge, represents a novel protein fold. The molecule is approximately 35 Å in length, 26 Å in width and 46 Å in height, allowing it to be embedded in the phospholipid membrane (Fig. 2c). Except for α A and α B, which interconnect TM2 with TM3 and TM7 with TM8, respectively, short loops link the transmembrane helices.

The ~100-residue ABI domain corresponds to TM4, TM5, TM6 and TM7 (Fig. 3a, c). These four transmembrane helices form an antiparallel helix bundle that has approximately two-fold symmetry and is surrounded by the less conserved helices TM1, TM2, TM3 and TM8 (Extended Data Fig. 3). The three conserved motifs of the ABI domain are positioned on TM4, TM5 and TM7 (Fig. 3a, c). Seven transmembrane helices delineate a large conical catalytic cavity (volume 1,400 Å³) (Fig. 2a, c and Fig. 3a). The catalytic residues E140, H173, H227 and N231 are located at the top of the cavity, inside the membrane (Figs 2c and 3a, b). This conformation of transmembrane helices allows unrestricted solvent

¹Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. ²Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ³JST, Research Acceleration Program, Membrane Protein Crystallography Project, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ⁴Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. ⁵Department of Life Sciences, Imperial College, London SW7 2AZ, UK. [†]Present addresses: Division of Biological Sciences, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India (K.K.); Cambridge Institute for Medical Research, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0XY, UK (R.B.D.).

*These authors contributed equally to this work.

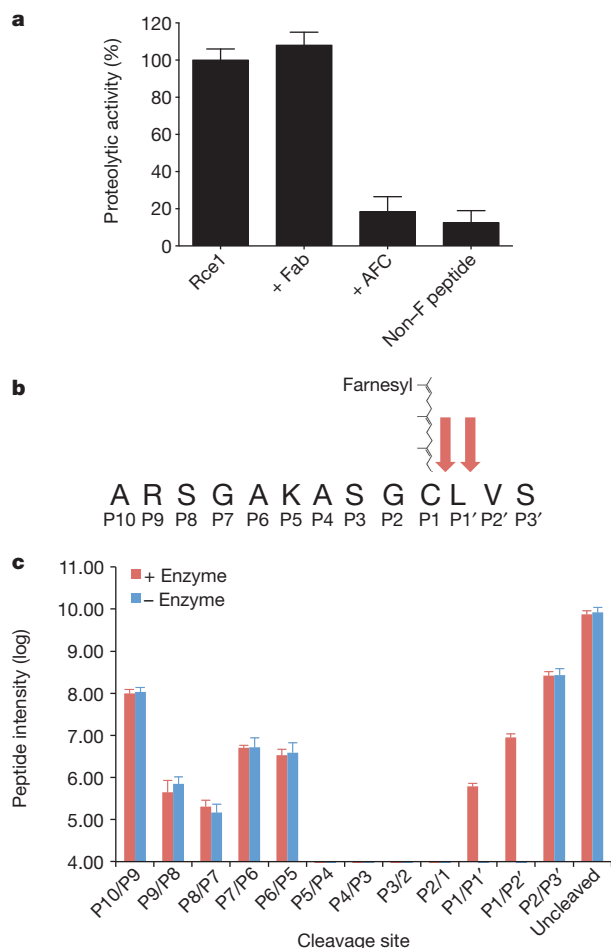


Figure 1 | MmRce1 is an endoprotease specific for farnesylated peptides.

a, The proteolytic activity of wild-type MmRce1 (Rce1; set to 100%) compared with the MmRce1–Fab complex (+ Fab), MmRce1 incubated with a molar excess of AFC (+ AFC) and MmRce1 incubated with a non-farnesylated peptide (+ Non-F peptide), as determined by fluorescence resonance energy transfer (FRET) assay. The data are presented as mean \pm s.d. for $n = 3$ experiments. **b**, A schematic representation of the RhoA-derived farnesylated peptide. The two cleavage sites identified by mass spectrometry are marked with red arrows. **c**, Semi-quantitative mass spectrometry data for the uncleaved and truncated farnesylated peptides (where, for example, P10/P9 denotes cleavage between P9 and P10 after incubation with MmRce1. Many of the truncated forms were also present in the no enzyme control sample (– Enzyme). These might be by-products of peptide synthesis that are isobaric with truncations corresponding to positions P2'/P3' and P6/P5 to P10/P9. Only the P1'/P2' and P1/P1' truncations (ARSGAKASGC(farnesyl)L and ARSGAKASGC(farnesyl), respectively) are found in the + Enzyme sample. The data are presented as the mean \pm 2 s.d. from $n = 4$ experiments.

access to the catalytic site from the cytoplasm at the base of the cavity. One side of the conical cavity is open to the membrane through a gap between TM2 and TM4. Access to the periplasm (the endoplasmic reticulum lumen) is blocked by the conserved residue R145, which interacts with T210 and the invariant E141 (Fig. 2a, b).

To understand the relevance of the MmRce1 structure in the context of the MmRce1–Fab complex, we assayed MmRce1 in the presence of the Fab. The formation of the MmRce1–Fab complex had no effect on the catalytic activity of MmRce1 (Fig. 1a and Extended Data Fig. 1), even though MmRce1 and the Fab form extensive contacts ($1,100 \text{ \AA}^2$) that would prevent conformational changes in the seven transmembrane helices that delineate the catalytic cavity. Thus, in the crystal structure of MmRce1, the catalytic residues are correctly aligned for cleaving a farnesylated peptide, although it is possible that conformational changes are required to accommodate larger, prenylated protein substrates. Both

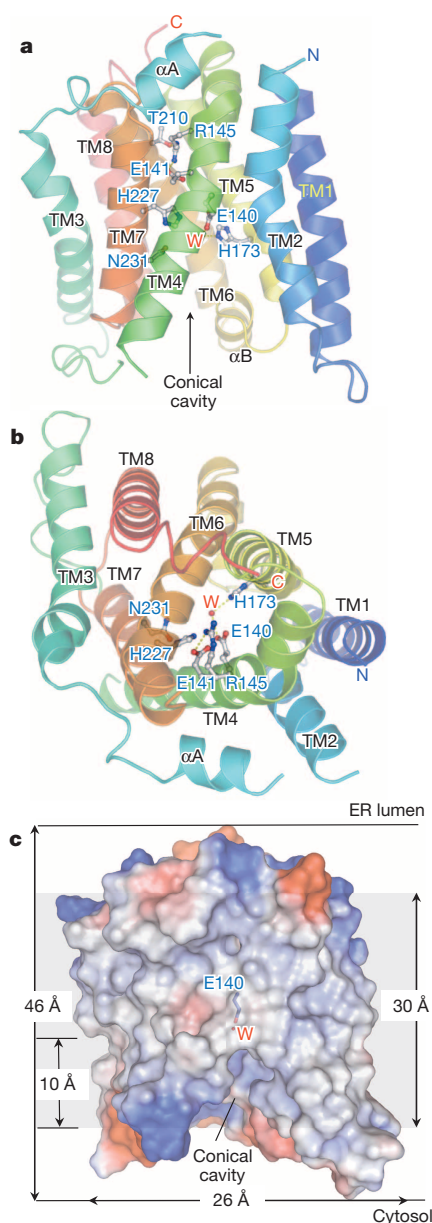


Figure 2 | MmRce1 is an IMP with eight transmembrane α -helices.

a, **b**, Two views showing ribbon representations of MmRce1 (molecule C from the asymmetric unit). The side chains of the five invariant ABI domain residues (and the conserved R145 and T210) are shown as ball-and-stick representations (with oxygen in red and nitrogen in blue). The catalytic water (W) molecule is shown as a red sphere. Hydrogen bonds are shown as dashed yellow lines. **a**, View of the molecule parallel to the membrane. **b**, View of the molecule from the luminal side of the endoplasmic reticulum (ER). **c**, The molecular surface of MmRce1, colour-coded by electrostatic potential ranging from blue (most positive) to red (most negative) to white (uncharged). The lipid membrane ($\sim 30 \text{ \AA}$) is indicated by the grey background, based on the distribution of nonpolar residues and transmembrane helices. The E140 side chain and the catalytic water molecule are shown. The catalytic water molecule is located $\sim 10 \text{ \AA}$ into the membrane.

the bacterial rhomboid GlpG^{24–27} and S2P⁸ have been proposed to undergo conformational shifts to mediate substrate gating, and structural changes in presenilin⁹ and FlaK¹² are necessary to align catalytic site residues.

Rce1 has no sequence similarity to other proteases¹³. It therefore represents a novel protease, and interestingly it has no paralogues in eukaryotes. Both cysteine-enzyme-based and metalloenzyme-based catalytic mechanisms have been proposed for Rce1 (refs 13, 15). However, the

absence of an evolutionarily conserved cysteine residue (Fig. 3c and Extended Data Fig. 3) precludes a thiol-based mechanism^{13,16}. Strong evidence also indicates that Rce1 is not a metalloenzyme. MmRce1 activity is unaffected by EDTA or Zn^{2+} (Extended Data Fig. 4a). Its concentration-dependent inactivation by 1,10-phenanthroline, a hydrophobic metal chelator, results from nonspecific protein unfolding (Extended Data Fig. 4). Furthermore, we did not detect Zn^{2+} ions bound to MmRce1 by using proton-induced X-ray emission (PIXE) or total reflection X-ray fluorescence (TXRF) (data not shown), and no Zn^{2+} ions were identified in the MmRce1 crystal structure.

All five conserved residues of the ABI domain (Fig. 3c) have been implicated in catalysis^{15,16}. Three residues of *S. cerevisiae* Rce1p—E156, H194 and H248 (equivalent to E140, H173 and H227 of MmRce1)—are critical for catalysis^{15,16}, whereas mutation of either E157 or N252 (E141 and N231 of MmRce1) impairs catalytic activity¹⁶ (see Fig. 3c). We confirmed that MmRce1 catalytic activity is strictly dependent on E140 and H173 and that mutation of either E141 or H227 severely disrupts activity (Fig. 3d). Similar to its eukaryotic homologues¹⁶, replacing the asparagine residue at position 231 with an alanine residue (N231A) abolished MmRce1 catalytic activity (Fig. 3d). Interestingly, in a few prokaryotic Rce1 homologues, aspartate is present instead of asparagine at position 231 (D231) (refs 13, 14) (Fig. 3c). When aspartate was substituted into MmRce1 at this position (N231D), the protease activity was ~40% of that of the wild-type protein (Fig. 3d). However, at a higher pH, the wild-type MmRce1 protease activity was unchanged,

whereas the N231D mutant was essentially inactive, suggesting that deprotonation of D231 inactivates MmRce1 (Fig. 3d). The low activity of the N231A mutant and the pH-dependent activity of MmRce1(N231D) indicate a catalytic role for a hydroxyl group at this position. Using thermal shift and circular dichroism assays, we detected no effect of the mutations on either MmRce1 stability or conformation (Extended Data Figs 5 and 6).

The buried position of E141 (Fig. 3a, b) suggests that it may have an indirect structural role and that the critical catalytic residues of MmRce1 are E140, H173, H227 and N231. The side chains of these four residues project into the solvent-filled catalytic cavity. E140 and H173, the residues that are most critical for catalysis, face each other from opposite transmembrane helices and coordinate a bridging water molecule (Fig. 3b). Two conserved aromatic residues (W144 and W169) contact the carboxylate side chain of E140, potentially raising its pK_a (Fig. 3b). Thus, a likely catalytic mechanism would involve E140 and H173 general base-catalysed deprotonation of a water molecule for nucleophilic attack on the scissile bond of the peptide substrate. The side chains of H227 and N231, positioned on successive helical turns of TM7 opposite the catalytic dyad of E140 and H173, are likely to donate hydrogen bonds to stabilize the oxyanion transition state. Protonation of the leaving amino group of the AAX tripeptide could be catalysed by either E140 or H173 (Extended Data Fig. 7).

Although Rce1 belongs to a novel protease family, the proposed catalytic mechanism has similarities to that of other proteases (Extended

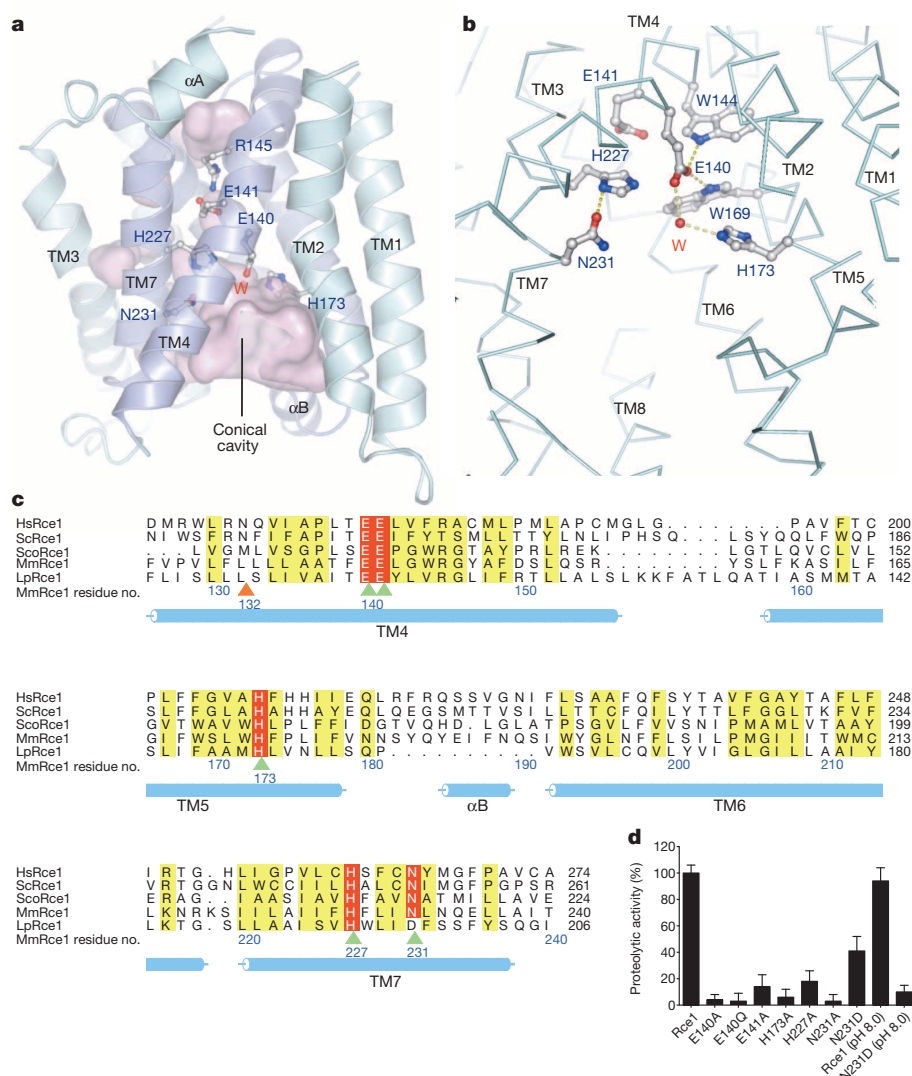


Figure 3 | The conserved catalytic ABI domain of MmRce1 is responsible for catalysis. **a**, Outline of the MmRce1 cavities as a surface representation (pink). The side chains of the five invariant ABI domain residues (and the conserved R145) are shown as ball-and-stick representations, and the catalytic water (W) is shown. TM4, TM5, TM6 and TM7 constitute the ABI domain (purple). **b**, Detailed interactions of the catalytic water molecule (red sphere) with the catalytic residues (E140 and H173). Hydrogen bonds are shown as dashed yellow lines. H227 and N231 are the residues that form the oxyanion hole. **c**, Multiple sequence alignment of the ABI domains from Rce1 homologues representing all three domains of life: *Homo sapiens* RCE1 (HsRce1; UniProt ID Q9Y256), *Saccharomyces cerevisiae* Rce1p (ScRce1; UniProt ID Q03530), *Streptomyces coelicolor* Rce1 (ScoRce1; UniProt ID Q9XAK4), *Methanococcus maripaludis* Rce1 (MmRce1; UniProt ID Q6LZY8) and *Lactobacillus plantarum* Rce1 (LpRce1; UniProt ID C6VK86). The residues that were mutated in this study and disrupt activity are indicated by green and orange arrows: the green arrows indicate catalytic residues, and the orange arrow indicates a putative farnesyl lipid-binding residue in TM4. Red indicates invariant residues, yellow indicates conserved residues. **d**, Proteolytic activity of wild-type MmRce1 (Rce1; set to 100%) and point mutants of MmRce1 towards a farnesylated peptide, as determined by FRET assay. Mutation of any of the five conserved ABI domain residues impairs MmRce1 catalytic activity. The data are presented as the mean \pm s.d. of $n = 3$ experiments.

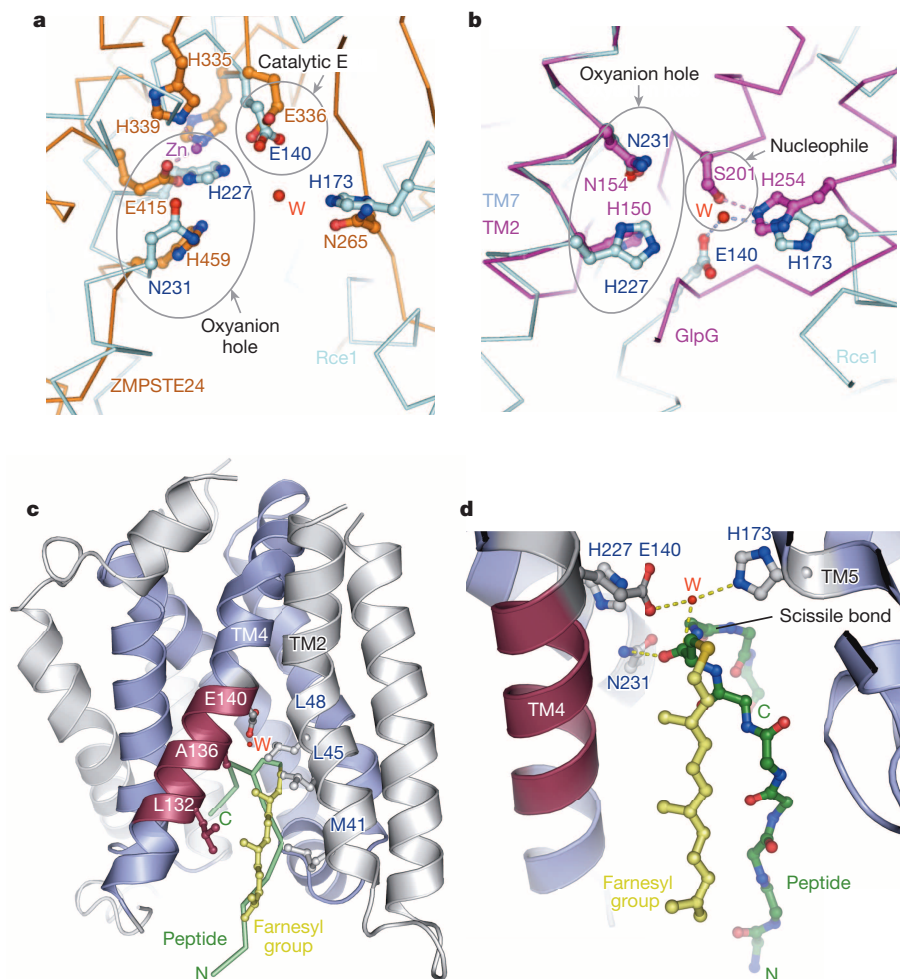


Figure 4 | MmRce1 catalytic mechanism has similarities to that of ZMPSTE24 and GlpG. **a**, Comparison of the catalytic site conformations of MmRce1 (cyan) and ZMPSTE24 (orange; PDB ID 4AW6)¹¹. Important side chains are shown as ball-and-stick representations, and the catalytic water molecule (W) is shown as a red sphere. The Zn atom is coloured pink and its coordination to Glu 415 is shown by a pink dashed line. Both proteins have similar cavities leading to the catalytic site. **b**, Comparison of the catalytic site conformations of MmRce1 (cyan) and GlpG (magenta; PDB ID 2O7L)²⁷. The catalytic water molecule from MmRce1 is located in a similar position to

the catalytic serine (S201) hydroxyl in GlpG. Hydrogen bonds are shown as dashed lines. **c**, Cartoon representation of MmRce1 with a farnesylated peptide (GAKASGC(farnesyl)LVS) modelled at its catalytic site. The ABI domain helices are shown in blue, and the conserved TM4 fragment (L132 to T138) is shown in dark pink. The nonpolar residues in TM2 and TM4 that flank the farnesyl lipid are labelled. **d**, Detailed view of the catalytic site of MmRce1 shown in **c**. The catalytic residues (E140 and H173), the catalytic water molecule and the oxyanion residues (H227 and N231) are shown. Hydrogen bonds are shown as dashed yellow lines.

Data Fig. 7). The membrane-associated zinc metalloproteinases S2P⁸, ZMPSTE24 and Ste24p (refs 10, 11) use a glutamate-activated water molecule for cleavage of the scissile bond (Fig. 4a and Extended Data Fig. 7), as does the recently described fungal glutamic peptidase SGP²⁸. In S2P, an asparagine residue is proposed to stabilize the oxyanion⁸. Our inspection of the ZMPSTE24 and Ste24p catalytic sites (refs 10, 11) suggests a similar role for asparagine and/or histidine residues (Fig. 4a and Extended Data Fig. 7). A striking similarity exists with the protease GlpG, in which the proposed oxyanion hole is formed by H150 and N154 in its invariant HX₃N motif, which is located on successive turns of a transmembrane helix^{25–27,29}, analogous to the proposed oxyanion hole of H227 and N231 in MmRce1. Superimposing H227 and N231 onto their equivalent residues in the HX₃N motif of GlpG reveals that the position of the proposed glutamate-activated nucleophilic water of MmRce1 exactly matches that of the nucleophilic hydroxyl group of the S201 side chain of GlpG (Fig. 4b). In support of the idea that H173 of MmRce1 activates the nucleophilic water for attack on the peptide substrate, the imidazole side chain of H173 superimposes with H254, in the serine–histidine catalytic dyad in GlpG (Fig. 4b and Extended Data Fig. 7). Thus MmRce1, analogous to metalloproteinases, aspartyl

proteases and SGP, uses a carboxylate to activate a nucleophilic water molecule; however, both MmRce1 and SGP are unlike other proteases in that they do not polarize the carbonyl group of the scissile peptide.

To understand how MmRce1 recognizes its substrates, we modelled a farnesylated peptide, based on the CAAX motif of RhoA, at the catalytic site of Rce1. The peptide adopts a β -hairpin conformation with the scissile bond positioned adjacent to the putative nucleophilic water molecule (Fig. 4c, d). The model is consistent with the proposed catalytic mechanism and suggests that the farnesyl lipid enters MmRce1's catalytic site from the membrane, sealing the opening between the non-polar faces of TM2 and TM4. The interaction of the farnesyl lipid at this site would position the isoprenyl cysteine relative to the catalytic water molecule, contributing to defining the CAAX motif cleavage site. There is sufficient space to accommodate the aliphatic residues of the CAAX motif adjacent to TM5, TM6 and TM7, whereas the six residues amino-terminal to the CAAX motif and the C-terminal X residue would occupy the solvent-filled conical cavity, which leads into the cytoplasm. The large volume of the cavity, which is capable of accommodating diverse residues, is consistent with the variation in the sequence of Rce1 substrates N-terminal to the CAAX motif. Mutations of TM5, which

might alter the size of the AAX-motif binding pocket, modify the specificity of yeast Rce1p for CAAX³⁰. In support of the model that the opening between TM2 and TM4 creates the farnesyl lipid-binding site, substitution of either L45 of TM2 or L132 of TM4 with a bulky tryptophan residue, to disrupt this site, substantially reduced MmRce1 activity (Extended Data Fig. 8), without affecting the protein structure (Extended Data Figs 5 and 6). To gain insight into the structure of eukaryotic Rce1 and its capacity to process both farnesylated and geranylgeranylated peptides, we generated a model of human RCE1 (Extended Data Fig. 8). Compared with MmRce1, the model predicts a longer TM4 helix that might provide a more extensive hydrophobic surface to promote favourable interactions with the C20 prenyl chain of a geranylgeranyl lipid.

This study establishes Rce1 as a founding member of a novel family of glutamate-dependent IMPs. Insights into the structure and mechanism of Rce1 have implications for the development of antagonists of CAAX motif processing, which have the potential to disrupt Ras signalling pathways.

METHODS SUMMARY

MmRce1 was expressed in *Escherichia coli* and purified to homogeneity by affinity chromatography and gel filtration. BALB/c mice were immunized with reconstituted MmRce1 proteoliposomes, and antibodies were raised to recognize conformational epitopes of the protein. MmRce1 was mixed with the Fab fragment Fab645-2, and the complex was crystallized as described in Methods. Diffraction data were collected on beamline I04-1 at the Diamond Light Source (UK). The structure was solved by molecular replacement, using an antibody Fab fragment structure (Protein Data Bank (PDB) ID 3VG9) as the search model. The data collection and refinement statistics are summarized in Extended Data Table 1. A fluorescence resonance energy transfer (FRET) assay was used to demonstrate the proteolytic activity of MmRce1 towards the C-terminal CAAX cleavage site of a farnesylated peptide modelled on human RhoA. MmRce1 was shown to be an endoprotease by mass spectrometry, and a farnesylated peptide was modelled into the catalytic site of the protein. The human RCE1 homologue was modelled in its apo state and as a complex with a geranylgeranylated peptide.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 May; accepted 11 October 2013.

Published online 1 December 2013.

1. Winter-Vann, A. M. & Casey, P. J. Post-prenylation-processing enzymes as new targets in oncogenesis. *Nature Rev. Cancer* **5**, 405–412 (2005).
2. Prior, I. A. & Hancock, J. F. Ras trafficking, localization and compartmentalized signalling. *Semin. Cell Dev. Biol.* **23**, 145–153 (2012).
3. Ahearn, I. M., Haigis, K., Bar-Sagi, D. & Philips, M. R. Regulating the regulator: post-translational modification of RAS. *Nature Rev. Mol. Cell Biol.* **13**, 39–51 (2012).
4. Boyartchuk, V. L., Ashby, M. N. & Rine, J. Modulation of Ras and a-factor function by carboxyl-terminal proteolysis. *Science* **275**, 1796–1800 (1997).
5. Schmidt, W. K., Tam, A., Fujimura-Kamada, K. & Michaelis, S. Endoplasmic reticulum membrane localization of Rce1p and Ste24p, yeast proteases involved in carboxyl-terminal CAAX protein processing and amino-terminal a-factor cleavage. *Proc. Natl Acad. Sci. USA* **95**, 11175–11180 (1998).
6. Michaelis, S. & Barrowman, J. Biogenesis of the *Saccharomyces cerevisiae* pheromone a-factor, from yeast mating to human disease. *Microbiol. Mol. Biol. Rev.* **76**, 626–651 (2012).
7. Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179–180 (2006).
8. Feng, L. *et al.* Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318**, 1608–1612 (2007).
9. Li, X. *et al.* Structure of a presenilin family intramembrane aspartate protease. *Nature* **493**, 56–61 (2013).
10. Pryor, E. E. Jr *et al.* Structure of the integral membrane protein CAAX protease Ste24p. *Science* **339**, 1600–1604 (2013).

11. Quigley, A. *et al.* The structural basis of ZMPSTE24-dependent laminopathies. *Science* **339**, 1604–1607 (2013).
12. Hu, J., Xue, Y., Lee, S. & Ha, Y. The crystal structure of GXGD membrane protease FlaK. *Nature* **475**, 528–531 (2011).
13. Pei, J. & Grishin, N. V. Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases. *Trends Biochem. Sci.* **26**, 275–277 (2001).
14. Kjos, M., Snipen, L., Salehian, Z., Nes, I. F. & Diep, D. B. The Abi proteins and their involvement in bacteriocin self-immunity. *J. Bacteriol.* **192**, 2068–2076 (2010).
15. Dolence, J. M., Steward, L. E., Dolence, E. K., Wong, D. H. & Poulter, C. D. Studies with recombinant *Saccharomyces cerevisiae* CaaX prenyl protease Rce1p. *Biochemistry* **39**, 4096–4104 (2000).
16. Plummer, L. J. *et al.* Mutational analysis of the Ras converting enzyme reveals a requirement for glutamate and histidine residues. *J. Biol. Chem.* **281**, 4596–4605 (2006).
17. Bergo, M. O. *et al.* Absence of the CAAX endoprotease Rce1: effects on cell growth and transformation. *Mol. Cell Biol.* **22**, 171–181 (2002).
18. Wahlstrom, A. M. *et al.* Rce1 deficiency accelerates the development of K-RAS-induced myeloproliferative disease. *Blood* **109**, 763–768 (2007).
19. Bergo, M. O. *et al.* On the physiological importance of endoproteolysis of CAAX proteins: heart-specific RCE1 knockout mice develop a lethal cardiomyopathy. *J. Biol. Chem.* **279**, 4729–4736 (2004).
20. Christiansen, J. R., Koldaivelu, S., Bergo, M. O. & Ramamurthy, V. RAS-converting enzyme 1-mediated endoproteolysis is required for trafficking of rod phosphodiesterase 6 to photoreceptor outer segments. *Proc. Natl Acad. Sci. USA* **108**, 8862–8866 (2011).
21. Otto, J. C., Kim, E., Young, S. G. & Casey, P. J. Cloning and characterization of a mammalian prenyl protein-specific protease. *J. Biol. Chem.* **274**, 8379–8382 (1999).
22. Hollander, I., Frommer, E. & Mallon, R. Human Ras-converting enzyme (hRCE1) endoproteolytic activity on K-Ras-derived peptides. *Anal. Biochem.* **286**, 129–137 (2000).
23. Hollander, I. J., Frommer, E., Aulabaugh, A. & Mallon, R. Human Ras converting enzyme endoproteolytic specificity at the P2' and P3' positions of K-Ras-derived peptides. *Biochim. Biophys. Acta* **1649**, 24–29 (2003).
24. Wu, Z. *et al.* Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nature Struct. Mol. Biol.* **13**, 1084–1091 (2006).
25. Baker, R. P., Young, K., Feng, L., Shi, Y. & Urban, S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl Acad. Sci. USA* **104**, 8257–8262 (2007).
26. Ben-Shem, A., Fass, D. & Bibi, E. Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl Acad. Sci. USA* **104**, 462–466 (2007).
27. Wang, Y. & Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proc. Natl Acad. Sci. USA* **104**, 2098–2102 (2007).
28. Fujinaga, M., Cherney, M. M., Oyama, H., Oda, K. & James, M. N. The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proc. Natl Acad. Sci. USA* **101**, 3364–3369 (2004).
29. Vinothkumar, K. R. *et al.* The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.* **29**, 3797–3809 (2010).
30. Trueblood, C. E. *et al.* The CaaX proteases, Afc1p and Rce1p, have overlapping but distinct substrate specificities. *Mol. Cell Biol.* **20**, 4381–4392 (2000).

Acknowledgements This work was funded by a Cancer Research UK grant to D.B. Part of this work was supported by the research acceleration program of the Japan Science and Technology agency and by BBSRC BB/G023425/1 (S.I.). We thank staff at the I04-1 Diamond Light Source for help with data collection, J. Yang for advice and discussions, I. De Moraes for support and T. Daviter for help with the circular dichroism experiments.

Author Contributions R.B.D. and I.M. cloned and established the purification protocols for MmRce1. I.M. purified MmRce1 and prepared MmRce1–Fab co-crystals. S.O. and S.I. developed MmRce1-specific monoclonal antibodies. I.M., K.K., N.C. and D.B. collected crystallographic data. K.K. and I.M. determined the MmRce1 structure. I.M. and Z.Z. generated MmRce1 mutants. I.M. performed the kinetic and stability assays. K.K. performed the protein modelling and peptide docking. A.J.T. and S.J.H. performed the mass spectrometry analysis. N.O'R. and G.B. synthesized geranylgeranylated peptides. I.M. and D.B. wrote the paper.

Author Information The coordinates for MmRce1 have been deposited in the Protein Data Bank database under accession number 4CAD. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (dbarford@mrc-lmb.cam.ac.uk).

Chromatin connectivity maps reveal dynamic promoter–enhancer long–range associations

Yubo Zhang^{1*†}, Chee-Hong Wong^{1*}, Ramon Y. Birnbaum^{2*†}, Guoliang Li^{3,4}, Rebecca Favaro⁵, Chew Yee Ngan¹, Joanne Lim⁴, Eunice Tai⁴, Huay Mei Poh⁴, Eleanor Wong⁴, Fabianus Hendriyan Mulawadi⁴, Wing-Kin Sung⁴, Silvia Nicolis⁵, Nadav Ahituv², Yijun Ruan³ & Chia-Lin Wei^{1,4}

In multicellular organisms, transcription regulation is one of the central mechanisms modelling lineage differentiation and cell-fate determination¹. Transcription requires dynamic chromatin configurations between promoters and their corresponding distal regulatory elements². It is believed that their communication occurs within large discrete foci of aggregated RNA polymerases termed transcription factories in three-dimensional nuclear space³. However, the dynamic nature of chromatin connectivity has not been characterized at the genome-wide level. Here, through a chromatin interaction analysis with paired-end tagging approach^{3–5} using an antibody that primarily recognizes the pre-initiation complexes of RNA polymerase II⁶, we explore the transcriptional interactomes of three mouse cells of progressive lineage commitment, including pluripotent embryonic stem cells⁷, neural stem cells⁸ and neurosphere stem/progenitor cells⁹. Our global chromatin connectivity maps reveal approximately 40,000 long-range interactions, suggest precise enhancer–promoter associations and delineate cell-type-specific

chromatin structures. Analysis of the complex regulatory repertoire shows that there are extensive colocalizations among promoters and distal-acting enhancers. Most of the enhancers associate with promoters located beyond their nearest active genes, indicating that the linear juxtaposition is not the only guiding principle driving enhancer target selection. Although promoter–enhancer interactions exhibit high cell-type specificity, promoters involved in interactions are found to be generally common and mostly active among different cells. Chromatin connectivity networks reveal that the pivotal genes of reprogramming functions are transcribed within physical proximity to each other in embryonic stem cells, linking chromatin architecture to coordinated gene expression. Our study sets the stage for the full-scale dissection of spatial and temporal genome structures and their roles in orchestrating development.

Gene regulatory networks are organized by spatial connectivity between distal regulatory elements (DREs) and their corresponding promoters³. Many of these DREs, including cell-specific enhancers, were

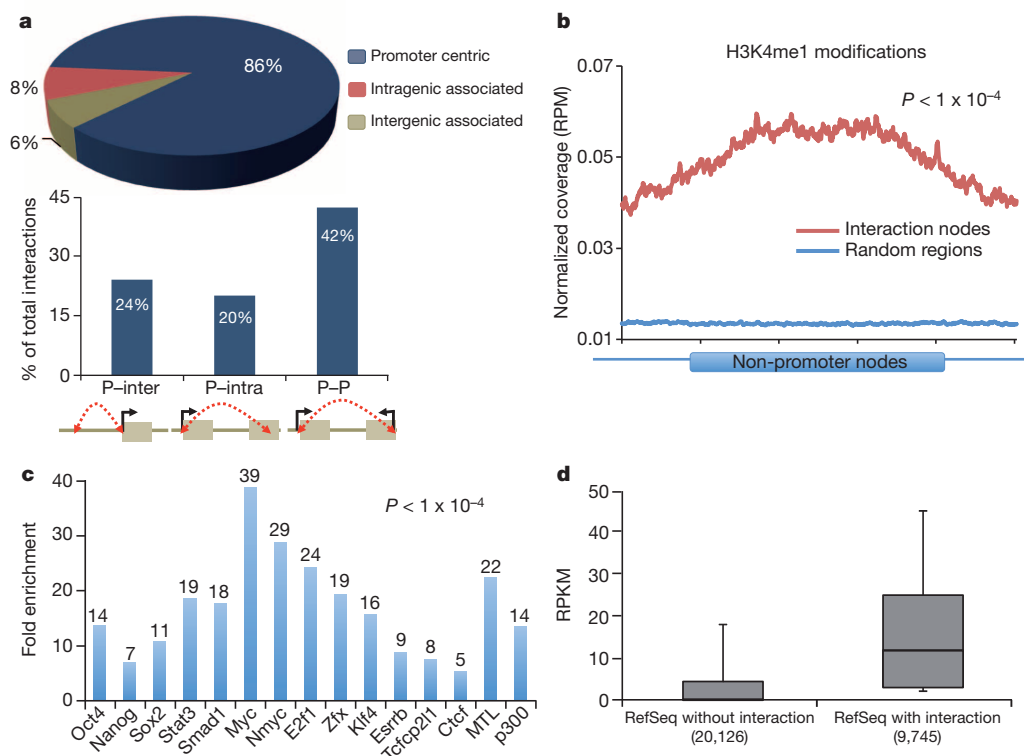


Figure 1 | RNAPII tethers promoter–enhancer interactions.

a, Distribution of promoter-, intragenic- and intergenic-centred interactions in ESCs. The percentages of P-inter, P-intra and P-P interactions are listed.

b, Enrichment of H3K4me1 signal density²⁸ (y axis: RPM, reads per million) along the non-promoter nodes (x axis). Signal from random control regions is shown in blue.

c, Enrichment of 12 transcription factors, Ctcf, p300 and MTL (multiple-transcription-factor-binding loci) occupancies at non-promoter regions. **d**, Boxplot expression (reads per kilobase per million reads (RPKM) as y axis) of the genes with versus without interactions (x axis) in ESCs.

¹Sequencing Technology Group, Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California 94598, USA. ²Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, UCSF, San Francisco, California 94158, USA. ³The Jackson Laboratory for Genomic Medicine, and Department of Genetic and Development Biology, University of Connecticut, 400 Farmington, Connecticut 06030, USA. ⁴Genome Institute of Singapore, 60 Biopolis Street, 138672 Singapore. ⁵Department of Biological Sciences and Biotechnology, University of Milano-Bicocca, 20126 Milano, Italy. [†]Present addresses: Department of Life Sciences, Faculty of Natural Sciences, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (R.Y.B.); National Heart, Lung, and Blood Institute, National Institutes of Health, Systems Biology Center, 9000 Rockville Pike, Bethesda, Maryland 20892, USA (Y.Z.).

*These authors contributed equally to this work.

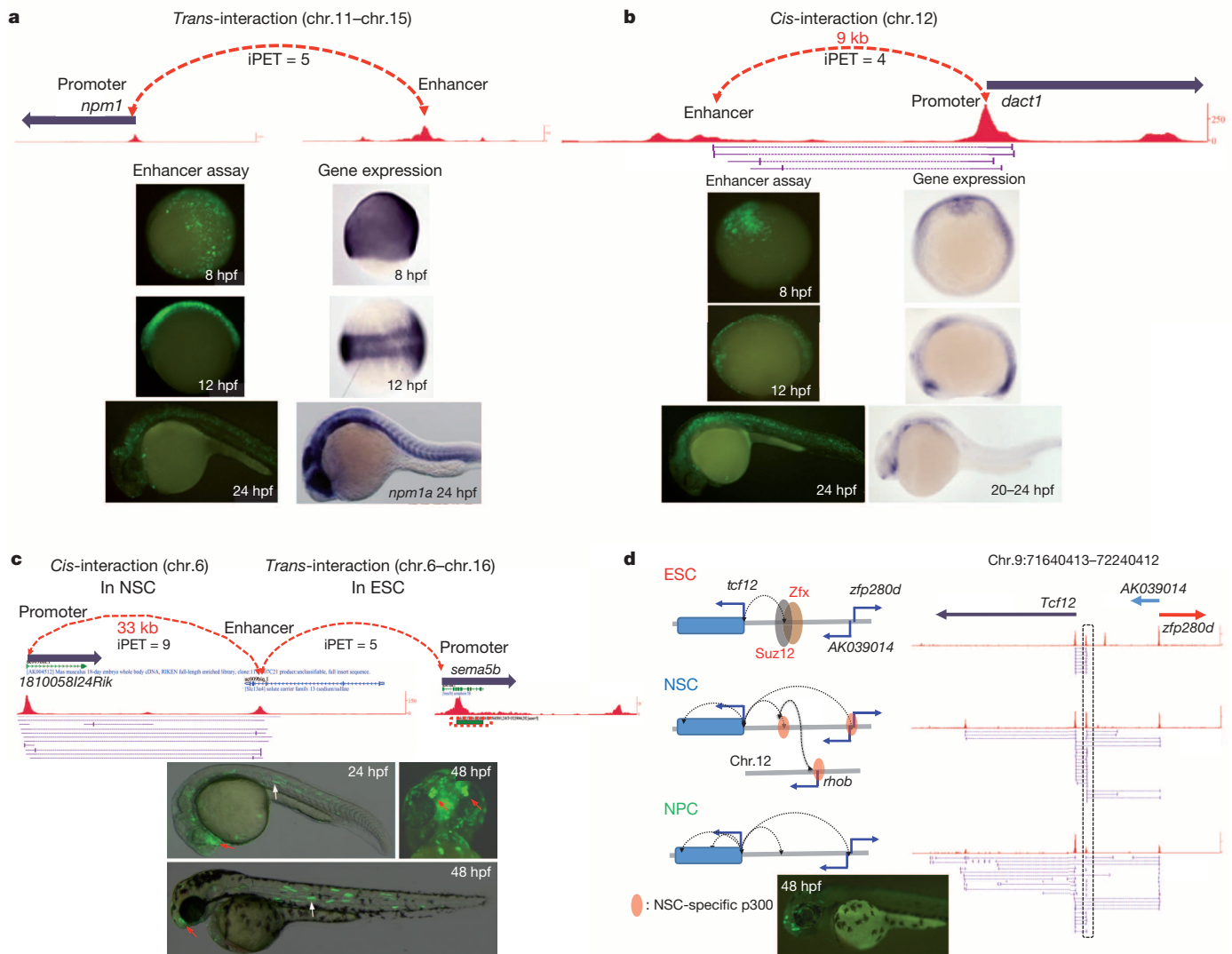


Figure 2 | Distal interaction loci display enhancer activities in zebrafish.

a, The PET-5 cluster showing an intergenic region, MTL12, on chromosome 15 that targets in *trans* the *npm1* promoter on chromosome 11 in ESCs. Embryos from 8, 12 and 24 hours post fertilization (hpf) are shown with the endogenous *npm1a* expression pattern as detected by whole-mount *in situ* hybridization (adapted from ref. 29 with permission). **b**, A *cis*-acting interaction region, MTL22, on chromosome 12, 9 kb upstream to the *dact1* target gene in ESCs. Zebrafish enhancer assays from 8, 12 and 24 hpf and endogenous *dact1* *in situ* hybridization are shown (adapted from ZFIN (<http://zfin.org/>) with

characterized for their vital function in development and differentiation². Increasing evidence has shown that DREs can function over long distances³, even on a different chromosome¹⁰ from their target genes. However, much of our current knowledge of cell-specific transcription regulation is based on extensive survey of DREs in the linear genome^{2,11,12}. The direct delineation of genome-wide DRE-promoter interactions is still very limited, and how chromatin structure regulates transcription is largely unknown.

To explore the promoter-associated chromatin interactomes, we used the RNA polymerase II (RNAPII) chromatin interaction analysis with paired-end tagging (ChIA-PET) approach³ (see Methods) in three murine cell types: embryonic stem cells (ESCs), neural stem cells (NSCs) and neurosphere stem/progenitor cells (NPCs). NSCs and NPCs are two widely used neural development models representing different commitment steps. NSCs are clonally derived early neural stem cells obtained by *in vitro* differentiation of ESCs⁸, whereas NPCs are neural progenitor cells derived *ex vivo* from the forebrain telencephalic hypon⁹. Further examination of their transcription profiles confirms

permission). **c**, A zebrafish-validated enhancer, mE12, located in the 13th intron of the *slc13a4* gene interacts with cell-specific target genes: 1810058124Rik (*cis* in NSCs) and *sema5b* (*trans* in ESCs). At 24 and 48 hpf, this enhancer drove GFP expression in the somitic muscles and olfactory vesicle.

d, Interactions mediated from the 'poised' enhancer in ESCs (top), NSCs (middle) and NPCs (bottom). The interactions are schematically depicted (left) and shown in the browser (right). Suz12, Zfx and p300 (ref. 12) binding are illustrated. Zebrafish embryo with representative forebrain expression at 48 hpf is shown at the bottom.

their expected cellular origins (Supplementary Information section 1). Chromatin immunoprecipitation (ChIP) efficiency and ChIA-PET library quality were evaluated (Extended Data Fig. 1 and Supplementary Information section 2) and the data were processed further to estimate reproducibility, noise and coverage (Supplementary Information sections 3, 4). Non-chimaeric, uniquely mapped PETs were used to define three classes of genome-wide information: the RNAPII-associated binding sites and long-range intra- and interchromosomal interaction clusters (Extended Data Fig. 2 and Supplementary Table 1). Using two independent approaches—quantitative PCR analysis of chromosome conformation capture (3C) and DNA fluorescent *in situ* hybridization (FISH)—we were able to validate the defined intra- and interchromosomal interactions (Supplementary Information section 5). In all we identified 40,000 RNAPII-bound interaction pairs present from a total of three cell types (Supplementary Table 2).

Consistent with its role in transcription initiation, this form of RNAPII, with a non-phosphorylated carboxy-terminal domain, tethers most (86–92%) of the chromatin interactions surrounding promoter regions

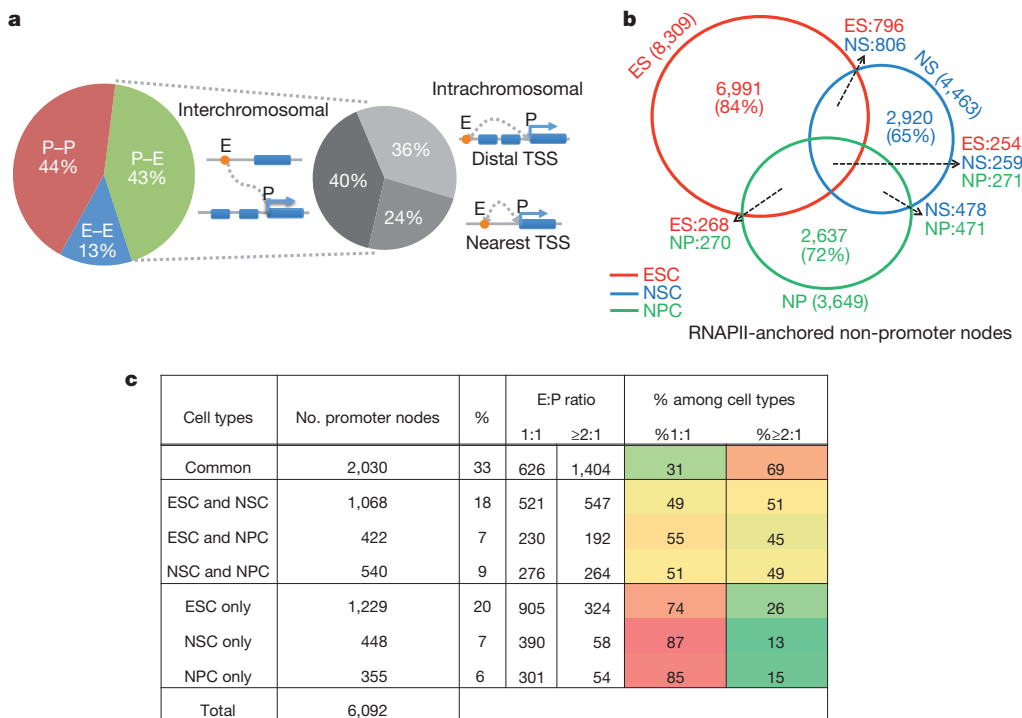


Figure 3 | Characterization of putative enhancer–promoter targeting. **a**, Distributions of P–P, E–E and E–P interactions in ESCs (left) and interchromosomal, intrachromosomal with nearest promoters and intrachromosomal with distal targeting promoters (right). The dotted lines indicate the relative portions of different targeting patterns derived from only the P–E interactions. **b**, Venn diagram of non-promoter nodes defined among ESCs, NSCs and NPCs. **c**, Putative enhancer–promoter pairing relations versus their cell specificities.

(± 2.5 kilobases (kb) of transcription start sites of UCSC known genes). In ESCs, roughly half of the promoter–tethered interactions connect two promoters (P–P), indicating the prevalence of promoter associations in the nucleus. The other half is distributed among promoters connecting to either intergenic (24%; P–inter) or intragenic (20%; P–intra) regions (Fig. 1a). Similar profiles are also found in NSCs and NPCs (Extended Data Fig. 3a). Through visual examination of these interactions, we identified many known enhancer–promoter (E–P) interactions, including the genic loci of *Oct4*, *Nanog*, *Phc1* and *Lefty1* (ref. 13) (Extended Data Fig. 4). Thus, these promoter-centric connectivity maps reveal large numbers of putative enhancers through both the inter- and intragenic interacting loci. In total, we identified 8,309, 4,463 and 3,649 putative DREs in ESCs, NSCs and NPCs, respectively. In ESCs, these distal loci exhibit enhancer characteristics; namely monomethylated histone H3 lysine 4 (H3K4me1) enrichment ($P < 1 \times 10^{-4}$), occupancy by pluripotent transcription factors and co-activator p300 (also known as Ep300), as well as sequence conservation¹⁴ (Fig. 1b, c and Extended Data Fig. 5). Furthermore, approximately half of the multiple-transcription-factor-binding loci (MTL) previously identified in ESCs¹² overlap with these distal interaction sites. These regions also share significant overlap ($P < 1 \times 10^{-4}$) with occupancy of mediator, a protein complex known to be associated with enhancers¹³. Last, the expression levels of the genes involved in the RNAPII interactions are significantly higher than those with no detected interaction ($P \leq 2.2 \times 10^{-16}$), suggesting that their promoters are transcriptionally more active (Fig. 1d and Extended Data Fig. 3b). Interestingly, among all the putative enhancers captured by RNAPII-bound interactions, we found 563 potential ‘poised enhancers’ connecting to the bivalent promoters¹⁵. Poised enhancers in ESCs are thought to ‘prime’ genes for subsequent for cell-type-dependent transcription activity during development¹⁶.

To directly confirm the *in vivo* enhancer activities of these non-promoter distal interaction loci, we carried out enhancer assays in transgenic zebrafish with green fluorescent protein (GFP) reporter genes¹⁷. Eleven out of 21 selected putative DRE loci drove reproducible and specific GFP expression patterns, indicating their spatiotemporal enhancer activity in zebrafish. (Fig. 2, Supplementary Table 3 and Supplementary Information section 6). Among them, a poised enhancer of *tcfl2*, a neuronal developmental gene, shows no activity at early

developmental stages but displays forebrain-specific activity at later developmental stages in zebrafish (Fig. 2d). Likewise in mouse, this locus is associated with the bivalent promoter, is bound by Suz12 and exhibits repressive trimethylated H3K27 (H3K27me3) modification in ESCs. This locus subsequently acquires p300 binding in NSCs and in correlation with high *Tcf12* expression (reads per kilobase per million reads (RPKM) 7, 119 and 75 in ESCs, NSCs and NPCs, respectively). Taken together, these distal interacting loci presumably not only function as developmental regulators but also retain their targeting specificities across different vertebrates.

These putative E–P interactions defined by ChIA-PET were used to decipher the nature of regulatory elements and their targeted gene associations in mouse stem cells. Being different from many previous studies that assumed proximity as the governing rule for E–P associations¹¹, our data suggest that a substantial fraction of these putative enhancers do not select their nearest promoters as their targets. In ESCs, 76% of the putative enhancer nodes interact beyond their closest active genes. Among them, 40% associated with interchromosomal interactions and 36% were involved in intrachromosomal interactions (Fig. 3a). Similar results were found in NSCs (77%) and NPCs (54%). Distal targeting has been reported in selected cases^{10,18}. Our data suggest that this could be a pervasive phenomenon throughout the genome. Besides long-range targeting, putative E–P interactions also exhibit high specificity for targeted genes and cell types. In all three cell types examined, 60–70% of the total RNAPII-tethered promoters are associated with only one distinct putative enhancer (E:P ratio = 1) and over 90% of the potential enhancer loci are associated with only one targeted gene (P:E ratio = 1) (Supplementary Table 4). Supplementary Tables 5 and 6 list the numbers of enhancers and promoters associated with each of their corresponding partner nodes in each cell type. Among the cell-specific interactions, the putative E–P interactions are the most prevalent type, suggesting that cell-specific genes targeted by their corresponding enhancers are the most distinctive features of the RNAPII-mediated chromatin interactions (Supplementary Table 7). For example, *Otx1* and *Meis2*, key forebrain-expressed homeobox genes, are connected with their cell-specific enhancers in NPCs, but not in NSCs. By contrast, NSCs contain specific E–P interactions involving genes expressed in the more posterior neuraxis, early development and in response to FGF2/EGF-like *Adam12*, *VAV3* and *Hoxa* (Extended Data

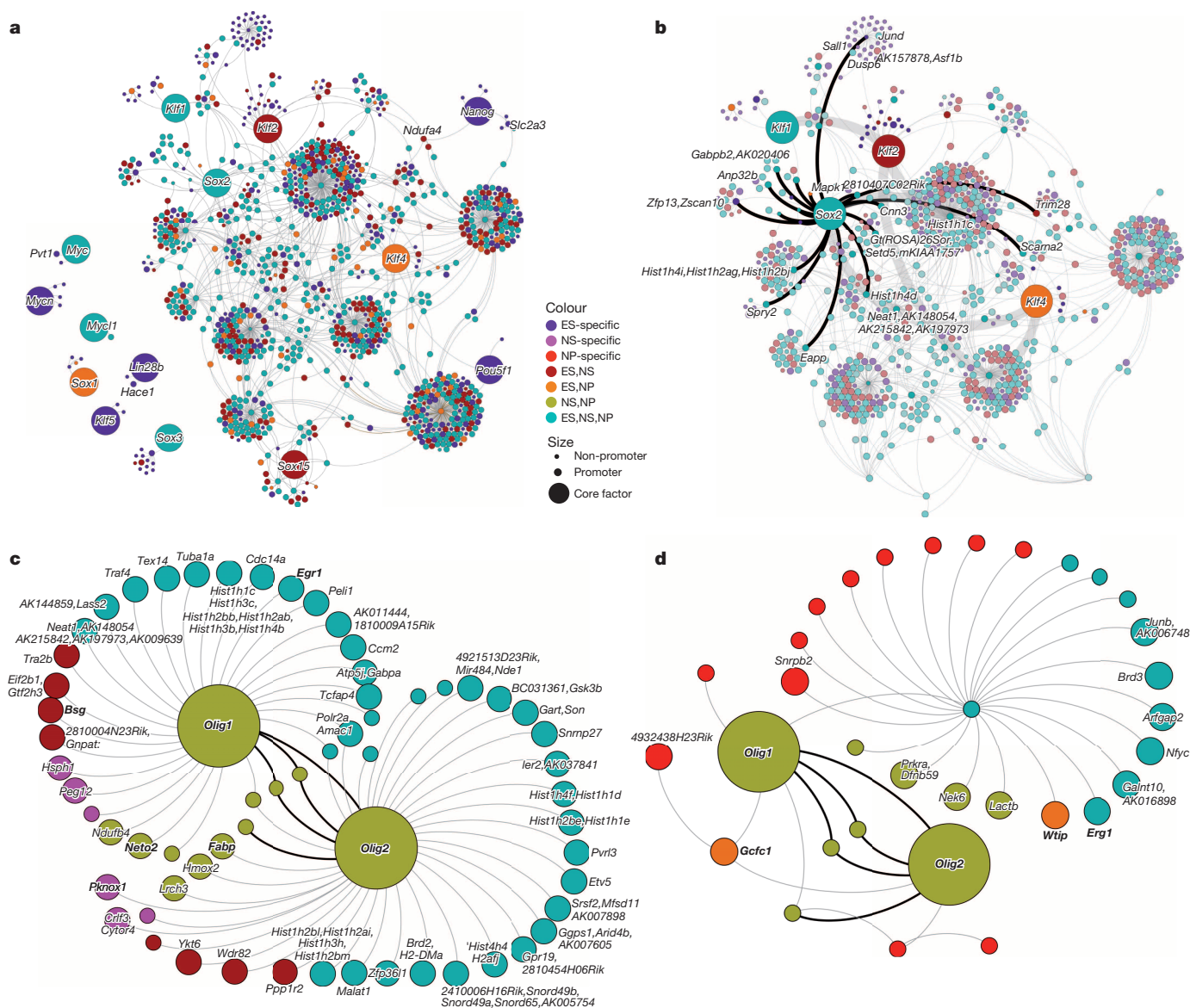


Figure 4 | Connectivity networks converged by key transcription regulator genes. **a**, Key re-programming gene network in ESCs. The connectivity was built through two hops of all interactions (light grey lines) mediated from 14 genes. Size of the circles represents the features of the nodes. Colours represent their cell specificities. **b**, *Sox2*-centric interaction map in ESCs. All of the

Fig. 6) are found. When all the interacting promoter and distal non-promoter nodes are compared separately, only a low percentage of the potential enhancers but a high percentage of the promoter nodes overlap among the three cell types (Fig. 3b, c), suggesting that the putative enhancers are mostly cell-type-specific but their targeted genes are largely common. When the cell-specific nature of these promoters was evaluated together with their interaction specificity, cell-specific promoters were found to be largely (74–87%) monogamous (E–P ratio of 1:1), whereas the universally expressed promoters are mostly (69%) promiscuous (E–P ratio of $\geq 2:1$) (Fig. 3c). Supplementary Table 8 lists the 6,092 interacting promoter nodes found in all three cell types, their cell specificity and enhancer connectivity, and examples are shown in Extended Data Fig. 7. Therefore, we speculate that distinct regulatory elements are used to target ubiquitously expressed genes in different cell types and such widespread differential enhancer usage not only functions to upregulate target genes but can also serve as a distinct mode to organize cell-specific transcription regulatory networks.

To evaluate whether transcription regulatory circuitry is reflected by the cell-specific chromatin organization, we explored the spatial chromatin

Sox2-directly-interacting genes are labelled. Thick grey lines highlight the connectivity between *Sox2*, *Klf1*, *Klf2* and *Klf4*. **c**, **d**, *Olig1*–*Olig2* interaction networks in NSCs (**c**) and NPCs (**d**). Thick black lines highlight their common interactions. Genes involved in neuronal developments are in bold.

connectivity of core reprogramming genes¹⁹ (*Pou5f1*, *Nanog*, *Lin28a*, *Sox2*, *Klf4* and *Myc*) in ESCs through the RNAPII interaction maps. The expression of these key reprogramming genes is known to govern pluripotency in ESCs through coordinated autoregulatory loops²⁰. Using connectivity analysis, three *Klf* genes (*Klf1*, *Klf2* and *Klf4*) were found to directly link to *Sox2* (Extended Data Fig. 8). When the network analysis was extended from one to two hops of connectivity, all of the reprogramming genes were found to be connected within one major hub, except for *Myc* and *Lin28a* (Fig. 4a), suggesting that they could be colocalized in the nucleus within the same ‘transcriptional factory’. Among them, *Nanog* and *Pou5f1* have limited connected edges whereas *Sox2* has extensive connectivity.

Sox2 can reprogram somatic cells to ESCs or NSCs²¹. In ESCs, the *Sox2* promoter connects to clusters of ESC-specific enhancers to other pluripotency related genes like *Sall1*, *Asf1b*, *Dusp6* and *Jund* (Fig. 4b). In NSCs, a very different *Sox2* connectivity profile is observed (Extended Data Fig. 9). Such cell-specific connectivity is mediated through differential enhancer usage. Similarly, distinct connectivity maps constructed by oligodendrocyte transcription factors *Olig1* and *Olig2*,

genes important for neural cell fate determination^{22,23}, are found in NSCs (Fig. 4c) and NPCs (Fig. 4d), respectively. As shown, *Olig1* and *Olig2* are directly connected to many genes critical for neuronal development in NSCs, including neuropilin (*Neto2*), *Fabp7* and *Bsg*.

RNAPII–chromatin interacting complexes captured here result from pre-initiation events and not all of them will proceed to active transcription or elongation²⁴. Therefore, the rich repertoires of *in vivo* chromatin interactions presumably comprise a mixture of promoters with various transcriptional activities. Using the wealth of other genome-wide transcription and epigenetic data sets, one can further discriminate the transcription states of the identified interacting promoters and derive the significance of different ‘transcription factories’.

Our current study illustrates the complexity and dynamics of the underlying chromatin structures in the nucleus; however, we recognize that not all the interactions are functional. With additional ChIA-PET analyses from various forms of RNAPII, deeper sequencing and increased coverage, we expect to capture more promoter-mediated interaction pairs and diverse types of regulatory elements. Before concluding new insights and principles, any functional model made solely from colocalization association has to be examined together with other integrated (epi)genomic information and requires extensive functional validations, like the zebrafish transgenic assays or genome editing tools like CRISPR/Cas²⁵. Finally, we expect that the recapitulation of such structure-based framework on a greater diversity of cell types and its integrated analyses shall further elucidate the mechanisms driving genome re-configuration, and to what extent it contributes to transcriptional regulation and cell specification.

METHODS SUMMARY

The ChIA-PET assay was performed using an RNAPII monoclonal antibody (8WG16, Covance) with E14 mouse ESC, NSC NS5 and NPC chromatin extracts. The sequence data were analysed as described²⁶. The interactome networks were constructed using non-overlapping genomic regions. Nodes were annotated using the UCSC Genes annotations as ‘Promoter’ if they were within 2.5 kb of any transcription start site. Interactomes from various gene sets were visualized in Gephi²⁷. ‘Parallel Force Atlas’ layout was used.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 December 2012; accepted 24 September 2013.

Published online 10 November 2013.

1. Neph, S. *et al.* Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150**, 1274–1286 (2012).
2. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
3. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
4. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet.* **43**, 630–638 (2011).
5. Fullwood, M. J. *et al.* An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
6. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).
7. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
8. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283 (2005).
9. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a β -geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).
10. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).

11. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
12. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
13. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
14. Hardison, R. C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nature Rev. Genet.* **13**, 469–483 (2012).
15. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
16. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
17. Li, Q. *et al.* A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Dev. Biol.* **337**, 484–495 (2010).
18. Nolis, I. K. *et al.* Transcription factors mediate long-range enhancer–promoter interactions. *Proc. Natl Acad. Sci. USA* **106**, 20222–20227 (2009).
19. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
20. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
21. Ring, K. L. *et al.* Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. *Cell Stem Cell* **11**, 100–109 (2012).
22. Pollard, S. M., Wallbank, R., Tomlinson, S., Grotewold, L. & Smith, A. Fibroblast growth factor induces a neural stem cell phenotype in foetal forebrain progenitors and during embryonic stem cell differentiation. *Mol. Cell. Neurosci.* **38**, 393–403 (2008).
23. Ligon, K. L. *et al.* Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. *Neuron* **53**, 503–517 (2007).
24. Levine, M. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 502–511 (2011).
25. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
26. Li, G. *et al.* ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**, R22 (2010).
27. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks <http://www.aiai.org/ocs/index.php/ICWSM/09/paper/view/154> and <https://gephi.org/users/publications/> (International AAAI Conference on Weblogs and Social Media, 2009).
28. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
29. Bolli, N. *et al.* Expression of the cytoplasmic NPM1 mutant (NPMc+) causes the expansion of hematopoietic cells in zebrafish. *Blood* **115**, 3329–3340 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank J. Mariani for the preparation of RNA from NPC; K. Murphy and A. Ku for their assistance with zebrafish enhancer assays; and A. Visel and A. Nord for discussion and their comments on the manuscript. S.N. and R.F. were supported by grants from ASTIL Regione Lombardia (SAL-19 ref. no. 16874), Telethon (GGP12152), Cariplo (Rif. 2010-0673) and AIRC (IG-5801). N.A. is supported by NINDS grant number R01NS079231, NICHD grant number R01HD059862, NHGRI grant numbers R01HG005058 and R01HG006768, NIDDK award number R01DK090382, NIGMS award number GM61390 and Simons Foundation SFARI no. 256769. R.Y.B. is supported by NINDS grant number R01NS079231 and the UCSF Program for Biomedical Breakthrough Research (PBBR). This work was supported by Agency for Science, Technology and Research (A*STAR), Singapore, the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231 and National Institutes of Health ENCODE grants (R01 HG004456-01, R01HG003521-01 and 1U54HG004557-01) to Y.R. and C.-L.W.

Author Contributions Y.Z. constructed ChIA-PET experiments and data evaluation. R.Y.B. and N.A. designed and performed zebrafish enhancer assays. C.-H.W. carried out data analysis. C.-Y.N. and Y.Z. performed the RNA-seq experiments. J.L. performed DNA-FISH. H.M.P., E.T., R.F. and E.W. prepared the cells and ChIP material. G.L., F.H.M., W.-K.S. and Y.R. designed the data processing pipeline. C.-L.W. and S.N. designed the experiments. Y.Z. and C.-L.W. wrote the paper. All authors provided intellectual input and approved the final manuscript.

Author Information All data described in this study have been deposited in the GEO under accession number GSE44067. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.-L.W. (cwei@lbl.gov).

Primary cilia are specialized calcium signalling organelles

Markus Delling^{1*}, Paul G. DeCaen^{1*}, Julia F. Doerner¹, Sebastien Febvay¹ & David E. Clapham^{1,2}

Primary cilia are solitary, non-motile extensions of the centriole found on nearly all nucleated eukaryotic cells between cell divisions. Only ~200–300 nm in diameter and a few micrometres long, they are separated from the cytoplasm by the ciliary neck and basal body. Often called sensory cilia, they are thought to receive chemical and mechanical stimuli and initiate specific cellular signal transduction pathways. When activated by a ligand, hedgehog pathway proteins, such as GLI2 and smoothened (SMO), translocate from the cell into the cilium^{1,2}. Mutations in primary ciliary proteins are associated with severe developmental defects³. The ionic conditions, permeability of the primary cilia membrane, and effectiveness of the diffusion barriers between the cilia and cell body are unknown. Here we show that cilia are a unique calcium compartment regulated by a heteromeric TRP channel, PKD1L1–PKD2L1, in mice and humans. In contrast to the hypothesis that polycystin (PKD) channels initiate changes in ciliary calcium that are conducted into the cytoplasm⁴, we show that changes in ciliary calcium concentration occur without substantially altering global cytoplasmic calcium. PKD1L1–PKD2L1 acts as a ciliary calcium channel controlling ciliary calcium concentration and thereby modifying SMO-activated GLI2 translocation and GLI1 expression.

We generated a transgenic ARL13B–EGFP mouse (*Arl13b-EGFP*^{tg}) in which primary and motile cilia show spectacular fluorescence labelling throughout the animal, developed a ratiometric Ca^{2+} sensor directed specifically to cilia, and patch-clamped individual cilia and measured the resting membrane potential and a calcium-permeant conductance (I_{cilia} ; see ref. 5). The *Arl13b-EGFP*^{tg} mice displayed ubiquitous green fluorescence only in primary and motile cilia but not in microvilli (Fig. 1, Extended Data Fig. 1 and Supplementary Video 1). After incubation of E14.5 embryos in clarifying ScaleA2 solution⁶ (Fig. 1a–c and Supplementary Video 2), three-dimensional imaging of ARL13B–GFP reveals its exclusive localization to cilia. Staining of wild-type and *Arl13b-EGFP*^{tg} retinal pigmented cell epithelia (mRPE) cells and mouse embryonic fibroblasts (MEFs) with ciliary markers confirmed that ARL13B–EGFP is localized exclusively to cilia without noticeable alteration of cilia morphology (Extended Data Fig. 2).

Primary cilia are not reliably loaded with Ca^{2+} -sensitive dyes, forcing experimentalists to rely on changes in cytoplasmic calcium concentration ($[\text{Ca}^{2+}]_{\text{cyto}}$) as an indirect indicator of ciliary calcium concentration ($[\text{Ca}^{2+}]_{\text{cilia}}$). To ameliorate this issue, we generated a genetically encoded calcium sensor that is targeted to the cilium by fusing GCaMP3 (ref. 7) to the carboxy terminus of smoothened (SMO–GCaMP3), enabling monitoring of $[\text{Ca}^{2+}]_{\text{cilia}}$ and $[\text{Ca}^{2+}]_{\text{cyto}}$ simultaneously. As shown in Extended Data Fig. 3 and Supplementary Video 3, addition of the Ca^{2+} ionophore, ionomycin, increased fluorescence in both the cytoplasm and cilium of hRPE1 cells stably expressing SMO–GCaMP3, although variation in ionomycin incorporation precludes precise $[\text{Ca}^{2+}]$ comparison between the two compartments.

To quantify $[\text{Ca}^{2+}]$ we developed a ratiometric SMO–mCherry–GCaMP3 calcium sensor⁸ (Fig. 2a, b). In hRPE1 cells stably expressing

SMO–mCherry–GCaMP3, the cilia-targeted GCaMP3 and mCherry fluorescence co-localized with the cilia-specific marker, acetylated tubulin (Extended Data Fig. 3e–h). To determine whether $[\text{Ca}^{2+}]_{\text{cilia}}$ could be increased without affecting $[\text{Ca}^{2+}]_{\text{cyto}}$, we ruptured the cilia membrane at the tip of the cilia (circle; Extended Data Fig. 3i, j) with an intense 1–2 s laser pulse (405 nm), leading to a rapid increase in $[\text{Ca}^{2+}]_{\text{cilia}}$ from the tip that travelled to the ciliary base (Supplementary Video 4). Peak $[\text{Ca}^{2+}]$ propagated down the cilia at a rate of $4.6 \pm 0.6 \mu\text{m s}^{-1}$, yielding an apparent diffusion constant (D_{Ca}) of $5.3 \mu\text{m}^2 \text{s}^{-1}$ (similar to $5\text{--}10 \mu\text{m}^2 \text{s}^{-1}$ for D_{Ca} in stellate cell dendrites⁹).

In order to closely monitor changes in $[\text{Ca}^{2+}]_{\text{cyto}}$ at the cilia–cytoplasm junction, we loaded the calcium indicator Fluo-4 into hRPE1 cells stably expressing SMO–mCherry–GCaMP3. Ciliary membrane rupture increased $[\text{Ca}^{2+}]_{\text{cilia}}$ and was detectable at the ciliary base after a ~40 s delay (Fig. 2c–e and Supplementary Video 5). More distant parts of the

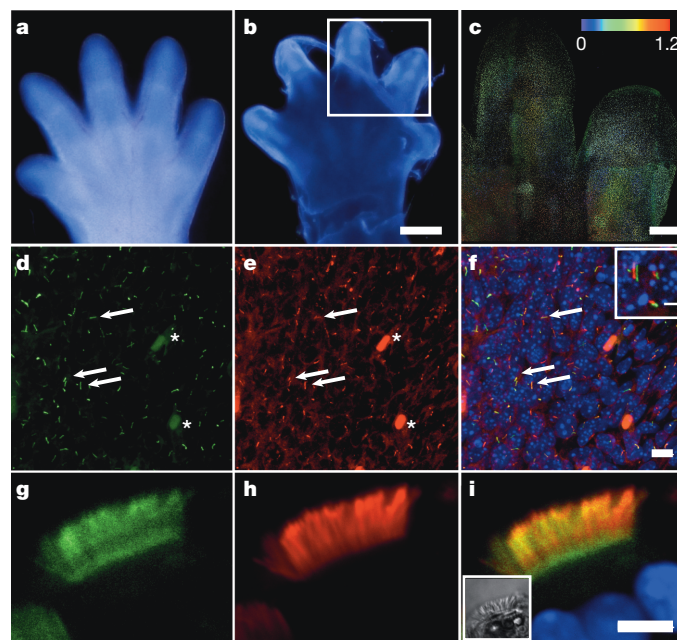


Figure 1 | ARL–GFP localizes to primary cilia and motile cilia *in vivo*. **a, b**, Fixed paw of E14.5 embryo (**a**), and after incubation in ScaleA2 (**b**). The inset indicates the area where z-stacks from 0 to 1.2-mm depth were acquired. Scale bar: 500 μm . **c**, ARL13B–EGFP expression in E14.5 paw. 3×4 z-stacks were stitched together, depth colour-coded and projected onto a plane. Scale bar: 200 μm . **d–f**, EGFP-positive cilia are present in fibroblasts within the digit (arrows) that overlap with adenyl cyclase III (ACIII) labelling in **e** and merged in **f**. Scale bar: 10 μm ; inset, 3 μm ; asterisks indicate autofluorescence. Red and green channels are offset to visualize co-localization. **g–i**, Motile cilia in the Fallopian tube express ARL13B–EGFP (**g**) that overlaps with acetylated tubulin (**h**). Overlay (**i**). Inset in **i** is DIC image. Scale bar: 5 μm .

¹Howard Hughes Medical Institute, Department of Cardiology, Boston Children's Hospital, 320 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

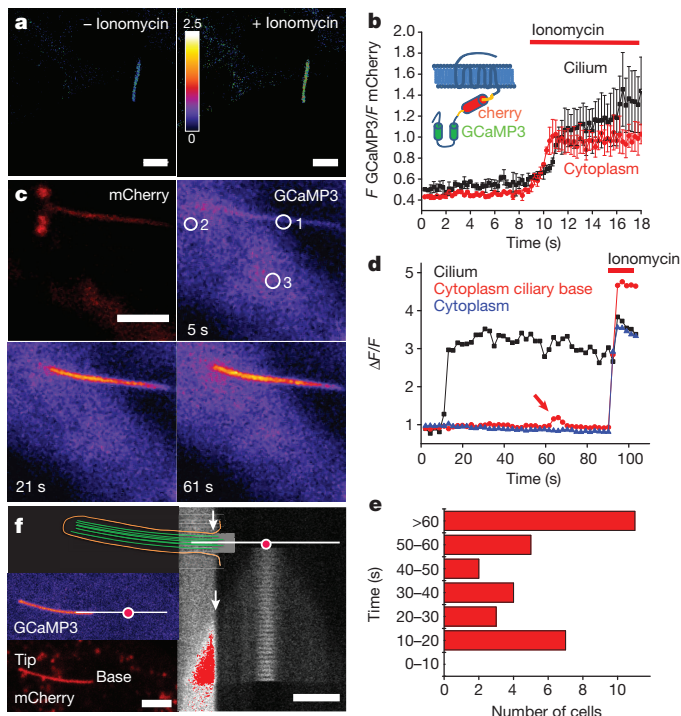


Figure 2 | SMO-mCherry-GCaMP3 localizes to the primary cilium and reports ciliary $[Ca^{2+}]$. $[Ca^{2+}]_{\text{cilia}}$ is poorly coupled to $[Ca^{2+}]_{\text{cyto}}$. **a**, hRPE1 cells expressing SMO-mCherry-GCaMP3 were treated with 5 μM ionomycin and changes in fluorescence were measured in the cilium and cytoplasm. **b**, Schematic of ratiometric calcium sensor and quantified fluorescence (F). **c**, **d**, After rupture of the ciliary tip with a laser pulse ($t = 10$ s), $[Ca^{2+}]_{\text{cilia}}$ rapidly increases, whereas $[Ca^{2+}]_{\text{cyto}}$ near the base of the cilium (circle position 2) increases only slightly 40 s later ($t = 61$ s). Ionomycin was added for normalization. Circles in **c** indicate areas where fluorescence was measured. $\Delta F/F$, changes in fluorescence. **e**, Distribution of lag times between ciliary and cytoplasmic $[Ca^{2+}]$ increases. **f**, Calcium diffusion from the cytoplasm to the cilium is not restricted. Red dot indicates the position of calcium uncaging whereas the white line indicates length and position of line scan. Arrows indicate cytoplasm to cilium transition. Scale bars: 5 μm . See also Supplementary Video 6.

cytoplasm remain unaffected by fluctuations at the base of the cilium. Elevation of $[Ca^{2+}]_{\text{cyto}}$ after ciliary rupture occurred infrequently and with varying delay; $\sim 40\%$ of the cells did not show any elevation in the cytoplasmic side of the cilia-cell junction even 60 s after $[Ca^{2+}]_{\text{cilia}}$ had saturated (Fig. 2e). Apparently, the Ca^{2+} sensor does not hinder the diffusion of Ca^{2+} from cilium to cytoplasm after rupture: in this case the cilium fills with 2 mM $[Ca^{2+}]$, which saturates the Ca^{2+} sensor. Next, we asked whether the protein-dense structure at the base of the cilium might act as a localized Ca^{2+} buffer and/or impermeant physical barrier to Ca^{2+} entry into the cytoplasm. We loaded SMO-mCherry-GCaMP3-expressing cells with the caged calcium chelator NP-EGTA-AM and uncaged Ca^{2+} in the cytoplasm in close proximity to the ciliary base. The velocity of this Ca^{2+} wave was $16 \pm 2 \mu\text{m s}^{-1}$ in the cytoplasm and continued at $15 \pm 2 \mu\text{m s}^{-1}$ in the cilium, indicating no significant delay of Ca^{2+} entering the cilium from the cytoplasm (Fig. 2f and Supplementary Video 6). In summary, because the ratio of cilium/cytoplasmic volume is exceedingly small (1:30,000), the tiny ciliary Ca^{2+} rivulet is rapidly diluted in the large cytoplasmic volume. Because the number of free calcium ions at 1 μM concentration within the ~ 0.5 fl cilioplasm is ~ 200 ions at an instant in time, ciliary $[Ca^{2+}]$ does not perturb cytoplasmic $[Ca^{2+}]$ substantially, nor does it initiate a measurable Ca^{2+} -induced Ca^{2+} release wave in the cytoplasm. These considerations led us to ask whether resting $[Ca^{2+}]_{\text{cilia}}$ differs from resting $[Ca^{2+}]_{\text{cyto}}$.

The dissociation constant (K_d) for the ratiometric ciliary Ca^{2+} sensor *in situ* was comparable to that in solution (625 nM versus 660 nM

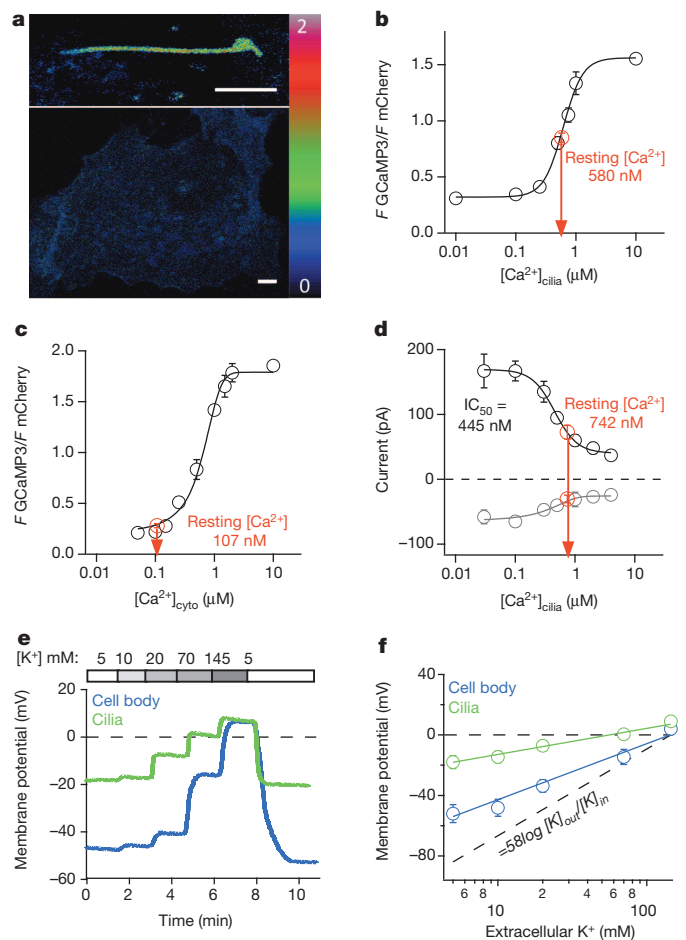


Figure 3 | Resting ciliary $[Ca^{2+}]$ is substantially higher than resting cytoplasmic $[Ca^{2+}]$. **a**, Live hRPE1 cell F $GCaMP3/F$ mCherry ratios in 2 mM $[Ca^{2+}]_e$ in the cilium (top) and cytoplasm (bottom). Scale bar: 5 μm . **b**, Plot of measured ciliary fluorescence ratio (average = 0.8 ± 0.12 , $n = 16$ cilia) versus estimated $[Ca^{2+}]_{\text{cilia}}$. Resting $[Ca^{2+}]_{\text{cilia}} = 580$ nM. **c**, Plot of measured cytoplasmic fluorescence ratio (average = 0.28 ± 0.02 , $n = 20$ cells) versus estimated $[Ca^{2+}]_{\text{cyto}}$. Resting $[Ca^{2+}]_{\text{cyto}}$ is 107 nM. **d**, By measuring current amplitudes in perforated patches (Methods), we estimated resting free $[Ca^{2+}]_{\text{cilia}}$ as 742 nM. Black circles indicate current at +100 mV; grey circles indicate current at -100 mV. **e**, Changes in cell (V_m) and cilia (V_{cilia}) potentials in response to external $[K^+]$. **f**, Average potential of the cell body and cilia plotted as a function of external $[K^+]$. V_m differs from V_{cilia} at all $[K^+]$ other than $[K^+]_e = 145$ mM; $P < 0.05$. The measured resting membrane potential is -18 mV for the cilia and -54 mV for the cell (\pm s.e.m., $n = 5$ cells and 4 cilia). The grey dashed line is the K^+ Nernst potential.

(ref. 7); Extended Data Fig. 4). In 2 mM extracellular Ca^{2+} concentration ($[Ca^{2+}]_e$), resting $[Ca^{2+}]_{\text{cilia}}$ was 580 nM in hRPE cells (Fig. 3b). We also calibrated the ratiometric Ca^{2+} sensor in the cytoplasm using cells that had not yet formed a cilium (SMO in the plasma membrane; Fig. 3a, bottom), and obtained a similar calibration curve and K_d (550 nM), suggesting that the sensor reported $[Ca^{2+}]$ similarly in cilia and cytoplasm. The average ratio for the sensor in the cytoplasm of 0.28 ± 0.02 corresponding to a resting $[Ca^{2+}]_{\text{cyto}}$ of < 110 nM, a normal resting cytoplasmic $[Ca^{2+}]^{10}$ (Fig. 3c). To test the surprisingly high resting $[Ca^{2+}]_{\text{cilia}}$ using an independent measurement, we patch-clamped SMO-EGFP-labelled cilia in the whole-cell configuration and varied free $[Ca^{2+}]_{\text{cilia}}$. High $[Ca^{2+}]_{\text{cilia}}$ (≥ 400 nM; $IC_{50} = 445$ nM (see Methods), Fig. 3d) inactivates I_{cilia} , enabling an independent calibration of $[Ca^{2+}]_{\text{cilia}}$. We then estimated the undisturbed resting $[Ca^{2+}]_{\text{cilia}}$ by comparing the current amplitude from perforated-patch measurements with the $[Ca^{2+}]$ calibration curve. The magnitude of I_{cilia} confirmed that $[Ca^{2+}]_{\text{cilia}}$ was ~ 7 -fold higher than resting $[Ca^{2+}]_{\text{cyto}}$ (Fig. 3c).

We next determined the cilia's resting membrane potential (E_{cilia}) by measuring changes in response to depolarizing concentrations of extracellular potassium ($[K^+]_e$) from hRPE1 SMO-EGFP cytoplasm and from its primary cilia. Consistently, E_{cilia} was >30 mV more positive than the cytoplasm ($E_{\text{cilia}} = -18 \pm 1$ mV; $E_{\text{cyto}} = -54 \pm 2$ mV, respectively; Fig. 3e) and significantly less $[K^+]_e$ (70 mM versus 129 mM) was required to depolarize the ciliary membrane potential to 0 mV (Fig. 3f). In summary, the cilium is a functionally distinct cell compartment with respect to ions. This calcium compartment is maintained by a favourable influx/efflux ratio: the large number of calcium-permeant channels⁵ or other ion channels/transporters can easily maintain high $[Ca^{2+}]$ in the small volume of the cilia, despite steady diffusion of Ca^{2+} into the cytoplasm at its base. An analogy is a water tower (Ca^{2+} within the cilia) connected to a large lake (cytoplasm) by a small pipe (basal body). Because ciliary $[Ca^{2+}]$ is high (~ 600 nM) compared to the cytoplasm (~ 100 nM), calcium flows from cilium to cytoplasm, but not cytoplasm to cilia. In addition, the approximately 30 mV gradient from cilia to cytoplasm further ensures asymmetry of Ca^{2+} flux. We speculate that these factors insulate cilia from the <1 μM Ca^{2+} fluctuations that characterize cytoplasmic Ca^{2+} signalling.

What are the consequences to cell function of ciliary ion independence from the cytoplasm? As we show in an accompanying paper⁵, I_{cilia} is mediated predominantly by the heteromeric channel PKD1L1–PKD2L1. *Pkd1l1*^{−/−} mice have a ciliary phenotype^{11,12}, but in this study we focused on the pore-forming subunit of I_{cilia} , PKD2L1. Although PKD2L1 was initially proposed to be a sour taste receptor, *Pkd2l1*^{−/−} mice display no sour taste deficit¹³. Thus, we tested *Pkd2l1*^{−/−} mice for a potential ciliary defect. Although we did not observe any major organ laterality defects, we observed intestinal malrotation in about 50% of the *Pkd2l1*^{−/−} knockout animals that was not observed in wild-type mice (Fig. 4a), indicating a mild penetrance of the phenotype. Intestinal malformations are associated with sonic hedgehog (SHH) pathway defects during early development^{14,15}.

Treatment of MEFs with the SMO agonist SAG directly activates the SMO receptor, leading to an upregulation of *Gli1* and *Ptch1* expression^{16,17}. As *Pkd2l1* is transcribed and localizes to cilia in wild-type MEFs (Fig. 4b and Extended Data Fig. 5), we next asked whether the SHH pathway might be affected in *Pkd2l1*^{−/−} mice by measuring the upregulation of GLI1 in response to stimulation with SAG. As shown in Fig. 4c, d, GLI1 protein increased from 1.9 ± 0.3 to 9.3 ± 0.7 AU (arbitrary units) in wild-type MEFs stimulated for 24 h with 400 nM SAG. In contrast, SAG upregulated GLI1 from 0.9 ± 0.04 to only 3.9 ± 0.6 in *Pkd2l1*^{−/−} embryonic MEFs. As shown in Fig. 4d, key members of the SHH pathway were not altered. However, *Arl13b-EGFP*^{tg} MEFs accumulated significant amounts of GLI2 at the distal tip of the cilium (Fig. 4e, f), as is required for activation of GLI2 and full activation of downstream transcription events¹⁸. In *Pkd2l1*^{−/−} \times *Arl13b-EGFP*^{tg} mutant cells, GLI2 accumulation at the ciliary tip was reduced by $\sim 50\%$ compared to wild-type cells. PKD2L1 is not required for cilia formation, as ciliary length and the percentage of ciliated cells did not differ significantly between genotypes (Fig. 4g, h). Finally, we asked whether SAG stimulation itself regulates I_{cilia} and thus ciliary $[Ca^{2+}]$. Although there was no immediate effect on ciliary current stimulation with SAG, after 24 h I_{cilia} amplitude and resting $[Ca^{2+}]_{\text{cilia}}$ increased in the cilium of wild-type MEFs (Extended Data Fig. 6). These data suggest that SAG initiates recruitment of PKD2L1 channels to the cilium rather than activating the channel.

Our results indicate that primary cilia are functionally distinct from the cytoplasm. I_{cilia} is encoded by a heteromeric PKD1L1–PKD2L1 channel and resting $[Ca^{2+}]_{\text{cilia}}$ is at least 0.4 μM higher than resting $[Ca^{2+}]_{\text{cyto}}$, which regulates trafficking of hedgehog-mediated transcription factors in the cilium. PKD2L1 is increased by SHH pathway stimulation, probably by channel recruitment to the cilium. Interestingly, the IFT25–IFT27 complex seems to be specific for transporting SHH components and an IFT25 mutant MEF showed an impaired GLI2 ciliary trafficking phenotype¹⁷ similar to *Pkd2l1* mutant MEFs. Because IFT25 has a unique

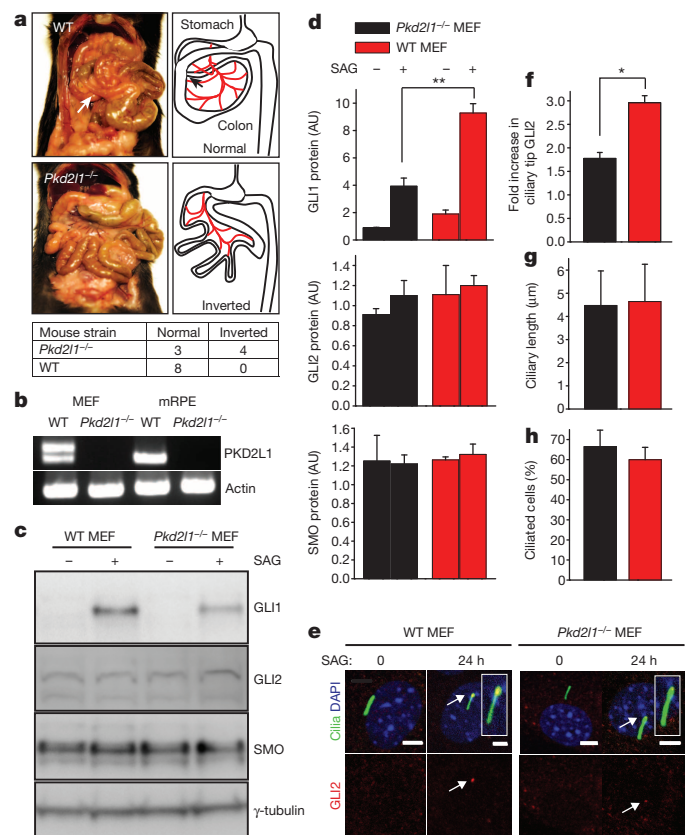


Figure 4 | *Pkd2l1*^{−/−} mice exhibit defects in SMO-mediated GLI1

activation. **a**, Intestinal malrotation in *Pkd2l1*^{−/−} mice, and comparison with wild type. Arrow indicates correct rotation of intestine. Schematic orientation of intestines is on the right; red indicates mesentery. **b**, RT-PCR of MEF and mRPE-derived cDNAs shows that PKD2L1 is expressed in both cell types. **c**, Western blot of GLI1, SMO, GLI2 and γ -tubulin expression in wild-type and *Pkd2l1*^{−/−} MEFs after stimulation for 24 h with 400 nM SAG. **d**, Top panel: GLI1 in wild-type (1.9 ± 0.3 versus 9.3 ± 0.7 after stimulation) and *Pkd2l1*^{−/−} MEFs (0.9 ± 0.04 versus 3.9 ± 0.6 after stimulation). $P < 0.01$. Middle panel: GLI2 in wild-type (0.9 ± 0.1 versus 1.1 ± 0.2 after stimulation) versus *Pkd2l1*^{−/−} MEFs (1.1 ± 0.3 versus 1.2 ± 0.1 after stimulation). Bottom panel: SMO in wild-type (1.3 ± 0.3 versus 1.2 ± 0.09 after stimulation) versus *Pkd2l1*^{−/−} MEFs (1.3 ± 0.04 versus 1.3 ± 0.1 after stimulation; $n = 3$ wild-type embryos, 5 *Pkd2l1*^{−/−} embryos). Error bars show \pm s.e.m. **e**, Localization of GLI2 at the distal tip of *Arl13b-EGFP*^{tg} and *Pkd2l1*^{−/−} \times *Arl13b-EGFP*^{tg} MEF cilia. Scale bar: 5 μm . Arrows point to GLI2 at cilia tips. Cilia are magnified in insets. **f**, Quantification of GLI2 protein accumulation at the ciliary tip after SAG stimulation; wild type = 3.0 ± 0.2 ($n = 40$ cilia each from 3 embryos) versus *Pkd2l1*^{−/−} MEFs = 1.8 ± 0.1 fold increase ($n = 40$ cilia each from 5 embryos). $P < 0.05$. **g**, **h**, Quantification of ciliary length (4.5 ± 1.5 μm for wild-type versus 4.6 ± 1.6 μm for *Pkd2l1*^{−/−} MEFs) and percentage of ciliated cells ($67 \pm 8\%$ for wild-type versus $60 \pm 6\%$ for *Pkd2l1*^{−/−} MEFs; $n = 40$ cilia each from 3 embryos).

Ca^{2+} binding site¹⁹, the observed SHH defect of *Pkd2l1* null mice is indicative that PKD2L1 may 'tune' $[Ca^{2+}]_{\text{cilia}}$ to optimize IFT25 function. $[Ca^{2+}]_{\text{cilia}}$ is probably also adjusted by other PKD members or even other as-yet-unidentified ion pumps or transporters in the cilium during development, which may explain the mild phenotype of the PKD2L1 mutant mouse compared to other SHH-deficient mutant mice (*Ift25*^{−/−} and *Shh*^{−/−} mice^{15,17}). An alternative, but not mutually exclusive, hypothesis is that acute I_{cilia} regulation by G-protein-coupled receptors and growth factors dynamically regulates ciliary trafficking. Further studies are necessary to determine whether other intermediates in the SHH pathway, such as SUFU–GLI dissociation or GLI proteolytic processing, are $[Ca^{2+}]$ -dependent. A second conclusion from this work is that the cilium funnels a small but steady Ca^{2+} load into the peri-ciliary cytoplasm.

METHODS SUMMARY

A ratiometric GCaMP3-based calcium sensor was expressed in cilia to measure ciliary calcium in hRPE1 cells. A transgenic mouse model expressing GFP in cilia was generated to measure ciliary calcium and membrane potential by patch-clamp recordings of cilia. Defects in SHH signalling of MEFs isolated from *Pkd2l1*^{-/-} mutant mice were quantified by western blotting and ciliary localization of GLI1 and GLI2 proteins.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 March; accepted 8 November 2013.

- Corbit, K. C. *et al.* Vertebrate Smoothed functions at the primary cilium. *Nature* **437**, 1018–1021 (2005).
- Huangfu, D. & Anderson, K. V. Cilia and Hedgehog responsiveness in the mouse. *Proc. Natl Acad. Sci. USA* **102**, 11325–11330 (2005).
- Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nature Rev. Genet.* **11**, 331–344 (2010).
- Praetorius, H. A. & Spring, K. R. Bending the MDCK cell primary cilium increases intracellular calcium. *J. Membr. Biol.* **184**, 71–79 (2001).
- DeCaen, P. G., Dellinger, M., Vien, T. N. & Clapham, D. E. Direct recording and molecular identification of the calcium channel of primary cilia. *Nature* <http://dx.doi.org/10.1038/nature12832> (this issue).
- Hama, H. *et al.* Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neurosci.* **14**, 1481–1488 (2011).
- Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
- Yasuda, R. *et al.* Imaging calcium concentration dynamics in small neuronal compartments. *Sci. STKE* **2004**, pl5 (2004).
- Soler-Llavina, G. J. & Sabatini, B. L. Synapse-specific plasticity and compartmentalized signaling in cerebellar stellate cells. *Nature Neurosci.* **9**, 798–806 (2006).
- Clapham, D. E. Calcium signaling. *Cell* **131**, 1047–1058 (2007).
- Vogel, P. *et al.* Situs inversus in *Dpdc/Poll*^{-/-}, *Nme7*^{-/-}, and *Pkd11l1*^{-/-} mice. *Vet. Pathol.* **47**, 120–131 (2010).
- Field, S. *et al.* Pkd11l1 establishes left-right asymmetry and physically interacts with Pkd2. *Development* **138**, 1131–1142 (2011).
- Horio, N. *et al.* Sour taste responses in mice lacking PKD channels. *PLoS ONE* **6**, e20007 (2011).
- Martin, V. & Shaw-Smith, C. Review of genetic factors in intestinal malrotation. *Pediatr. Surg. Int.* **26**, 769–781 (2010).
- Ramalho-Santos, M., Melton, D. A. & McMahon, A. P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763–2772 (2000).
- Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* **24**, 670–682 (2010).
- Keady, B. T. *et al.* IFT25 links the signal-dependent movement of Hedgehog components to intraflagellar transport. *Dev. Cell* **22**, 940–951 (2012).
- Kim, J., Kato, M. & Beachy, P. A. Gli2 trafficking links Hedgehog-dependent activation of Smoothed in the primary cilium to transcriptional activation in the nucleus. *Proc. Natl Acad. Sci. USA* **106**, 21666–21671 (2009).
- Bhogaraju, S., Taschner, M., Morawetz, M., Basquin, C. & Lorentzen, E. Crystal structure of the intraflagellar transport complex 25/27. *EMBO J.* **30**, 1907–1918 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Mouse Gene Manipulation Facility of the Boston Children's Hospital Intellectual and Developmental Disabilities Research Center (IDDR; NIH P30-HD 18655). We thank the Image and Data Analysis Core (IDAC) at Harvard Medical School for help with three-dimensional reconstruction. We also thank A. Salic for anti-SMO and anti-GLI2 antibodies. P.G.D. was supported by NIH T32-HL007572. We thank M. Desai for graphical assistance and A. von Gise and H. Tukachinsky (Harvard Medical School) and the members of the Clapham laboratory for advice and discussion.

Author Contributions All authors designed or conducted experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.C. (dclapham@enders.tch.harvard.edu).

Direct recording and molecular identification of the calcium channel of primary cilia

Paul G. DeCaen^{1*}, Markus Delling^{1*}, Thuy N. Vien² & David E. Clapham^{1,3}

A primary cilium is a solitary, slender, non-motile protuberance of structured microtubules (9+0) enclosed by plasma membrane¹. Housing components of the cell division apparatus between cell divisions, primary cilia also serve as specialized compartments for calcium signalling² and hedgehog signalling pathways³. Specialized sensory cilia such as retinal photoreceptors and olfactory cilia use diverse ion channels^{4–7}. An ion current has been measured from primary cilia of kidney cells⁸, but the responsible genes have not been identified. The polycystin proteins (PC and PKD), identified in linkage studies of polycystic kidney disease⁹, are candidate channels divided into two structural classes: 11-transmembrane proteins (PKD1, PKD1L1 and PKD1L2) remarkable for a large extracellular amino terminus of putative cell adhesion domains and a G-protein-coupled receptor proteolytic site, and the 6-transmembrane channel proteins (PKD2, PKD2L1 and PKD2L2; TRPPs). Evidence indicates that the PKD1 proteins associate with the PKD2 proteins via coiled-coil domains^{10–12}. Here we use a transgenic mouse in which only cilia express a fluorophore and use it to record directly from primary cilia, and demonstrate that PKD1L1 and PKD2L1 form ion channels at high densities in several cell types. In conjunction with an accompanying manuscript², we show that the PKD1L1–PKD2L1 heteromeric channel establishes the cilia as a unique calcium compartment within cells that modulates established hedgehog pathways.

Patch clamp of primary cilia is challenging due to their small dimensions (~0.2–0.5 μm in width, 1–12 μm in length), making them difficult to identify *in vivo*. Using a human retina pigmented epithelium cell line stably expressing the cilia-specific enhanced green fluorescent protein (EGFP)-tagged smoothened gene (hRPE SMO–EGFP), the cilia could be visualized under confocal fluorescence microscopy and recorded using the method we describe here: whole-cilia patch clamp (Fig. 1a, Extended Data Fig. 1a and Supplementary Video 1). After establishing >16 G Ω seals and rupturing the cilia membrane, we recorded a surprisingly large, outwardly rectifying, non-inactivating current (I_{cilia}). Notably, I_{cilia} was recorded from cilia attached or detached from the cell body (Fig. 1b, c and Supplementary Video 2). Current density measured in the detached cilia patch was 56-fold higher than that measured from the hRPE cell body (Methods). These measurements indicate that the primary cilium is partly insulated from the cell body by the structures at the cell–cilia junction (Extended Data Fig. 1a). The outwardly rectifying current was cation-non-selective (Fig. 1d) with relative permeabilities of $\text{Ca}^{2+} \approx \text{Ba}^{2+} > \text{Na}^+ \approx \text{K}^+ > \text{NMDG}$ (Extended Data Fig. 1b).

Consistent with the whole-cilia currents, single-channel amplitudes were outwardly rectifying (Fig. 1e) and mean open times substantially longer at more depolarized potentials. Extracellular uridine and adenosine phosphates (UDP, ADP, ATP) activated the ciliary current in perforated-cilia recordings, whereas the non-selective antagonists Gd^{3+} and ruthenium red blocked it (Extended Data Fig. 1c–e). Several cell-permeable calmodulin antagonists also activated the conductance. On the basis of the dimensions of the cilia, we estimate the average membrane surface area to be ~6.3 μm^2 (~0.063 pF). Assuming the entire outward rectifying

current is carried by the 96-pS conductance, we estimate the primary cilia channel density to be 128 ± 13 channels μm^{-2} , similar to endogenous channel densities calculated from excitable tissue plasma membranes

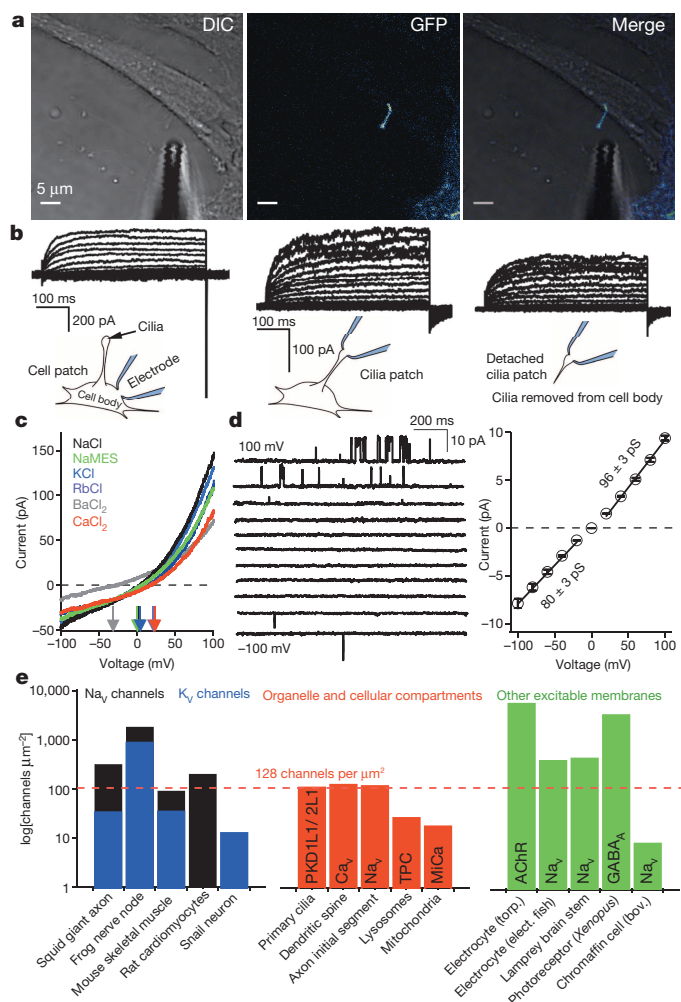


Figure 1 | A calcium-selective ion channel is richly expressed in primary cilia. **a**, Confocal image of an hRPE SMO–EGFP cell and patch-clamp electrode. **b**, Whole-cell leak-subtracted currents elicited by 1-s depolarizing pulses from -100 to 100 mV in +5-mV increments recorded from the cell body, primary cilia and an excised primary cilia (recorded from the same cilium). **c**, Whole-cell currents activated by ramp voltage protocols from -100 to +100 mV measured from the primary cilia where extracellular Na^+ -based saline was replaced by the cation indicated. **d**, Single-channel currents activated by 1.5-s depolarizations to the indicated potentials (left) and average current amplitudes (right; \pm s.e.m., $n = 8$ cilia). **e**, Estimated endogenous cilia ion channel densities compared to those from other biological preparations^{17,27–30}.

¹Howard Hughes Medical Institute, Department of Cardiology, Children's Hospital Boston, 320 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Neuroscience, Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, Massachusetts 02111, USA. ³Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

and larger than those found in intracellular compartments (Fig. 1f). Thus, the primary cilium is richly populated with Ca^{2+} -permeant, relatively non-selective cation channels that enable a much higher dynamic range of ciliary Ca^{2+} concentration compared to the cytoplasm.

We generated a transgenic mouse expressing the cilia-specific *Arl13b* gene carboxy-terminally tagged with EGFP (*Arl13b-EGFP^{tg}* mouse)² and isolated primary cells from mouse RPE (mRPE) and embryonic fibroblasts (MEFs). Primary cilia currents from these primary cells were outwardly rectifying with the same conductance and pharmacological properties as observed in the cilia from the human RPE cell line (Fig. 2 and Extended Data Fig. 2). In addition, I_{cilia} was observed in the cilia of a human kidney-derived inner medullary collecting duct cell line (IMCD) stably expressing ARL-EGFP (Fig. 2c and Extended Data Fig. 2d). Ciliary single-channel conductances were identical in all four cell types (Fig. 2e), and were activated by extracellular ATP and blocked by Gd^{3+} in perforated patch recordings. ATP addition to the bath significantly increased the probability of channel opening (P_o) and mean open times (5–7-fold, Fig. 2f and Extended Data Fig. 2e). Because these were in the 'on-cilia' patch configuration (bath-applied ATP), we reasoned that ATP binds a G-protein-coupled purinergic receptor to initiate activation of the channels in the patch. Thus, I_{cilia} is a common feature of many cell types.

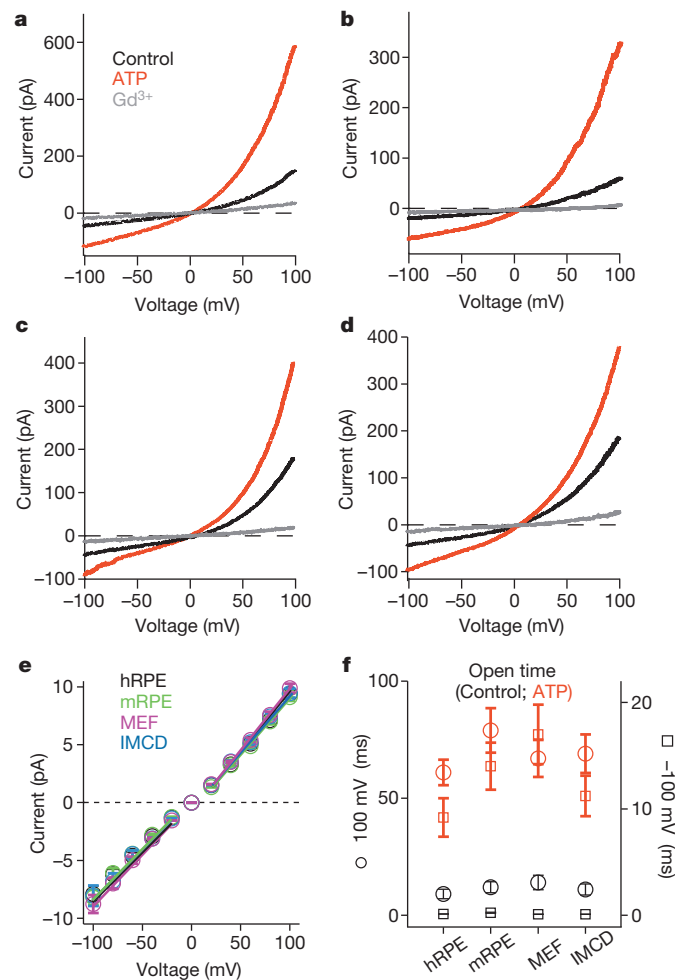


Figure 2 | Primary cilia currents measured from four different cell types. **a–d**, Averaged cilia current traces in control and bath-applied 100 μM ATP or 10 μM Gd^{3+} from: human RPE cell line stably expressing smoothened-EGFP (**a**); primary mRPE cells from the *Arl13b-EGFP^{tg}* mouse (**b**); kidney IMCD cell line stably expressing ARL-EGFP (**c**); and primary embryonic fibroblasts from the *Arl13b-EGFP^{tg}* mouse (**d**). **e**, Average single-channel current–voltage relation. The slope is used to estimate conductance (\pm s.e.m., $n = 4$ –7 cilia). **f**, Average open times in the presence and absence of ATP at -100 and $+100$ mV potentials measured from the cilia of RPE SMO-EGFP cells (\pm s.e.m., $n = 6$ cilia).

Analysis of hRPE transcripts confirmed expression of several purported ciliary channels^{13–16}, including TMC7, TRPV4 and PKD1, PKD2, PKD1L1 and PKD2L1 (Extended Data Fig. 3a). Only short interfering (si)RNAs specific for PKD1L1 and PKD2L1 reduced both inward and outward currents (Extended Data Fig. 3c–d). These results indicate that I_{cilia} is conducted by either PKD1L1 or PKD2L1 independently, or together as a heteromeric ion channel. To verify I_{cilia} channel proteins, we patch clamped cilia of homozygous PKD2L1 knockout (*Pkd2l1*^{−/−}) crossed with *Arl13b-EGFP* mice. The much-reduced *Pkd2l1*^{−/−} MEF ciliary current was linear and failed to activate when stimulated by calmidazolium (Fig. 3b, c). These data establish that PKD2L1 is a component of I_{cilia} .

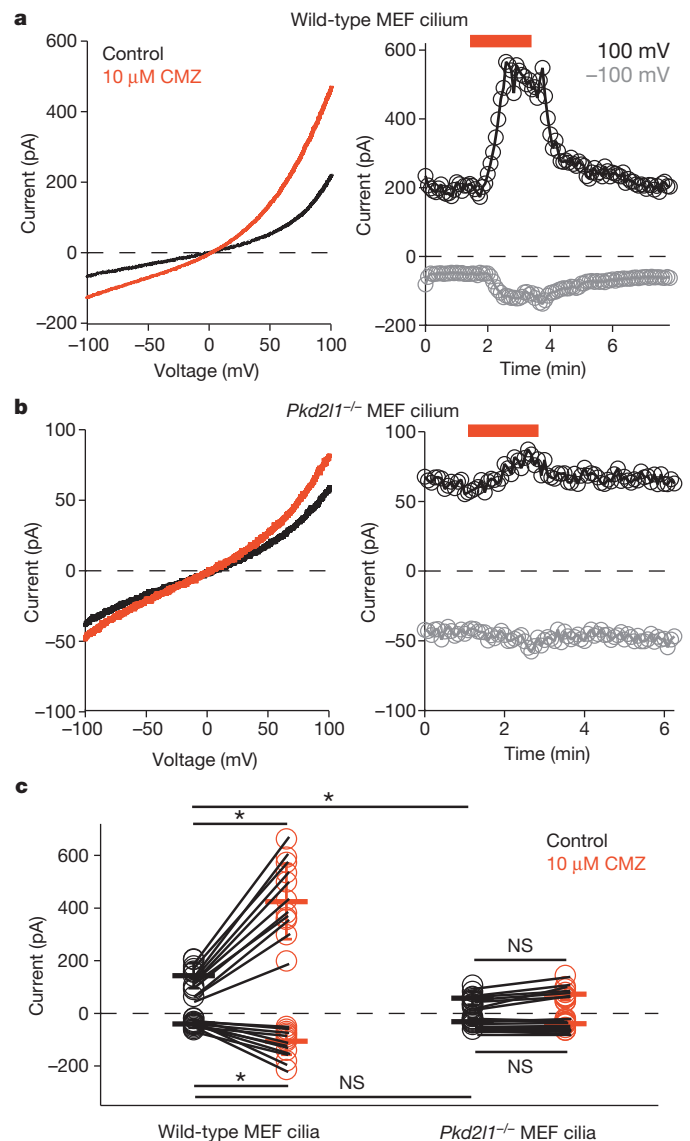


Figure 3 | MEF primary cilium currents compared from wild-type and *Pkd2l1* null animals. **a, b**, Left: cilia currents recorded from MEFs isolated from wild-type (**a**) and *Pkd2l1*^{−/−} (**b**) mice. Currents were elicited by a series of ramps from -100 to $+100$ mV in control conditions (black traces) or in the presence of 10 μM calmidazolium (CMZ, red trace). Right: resulting current amplitudes (-100 mV, grey circles; $+100$ mV, black circles) plotted as a function of time. Red bar indicates application of extracellular 10 μM CMZ. **c**, Scatter and whisker (\pm s.d.) plots of current magnitudes at $+100$ mV and -100 mV from MEF cilia. Individual cilia are represented as connected circles in control (black) and after calmidazolium (red). Averages are indicated by the dark horizontal lines. Student's *t*-test results: * $P < 0.05$; NS, $P > 0.05$; $n = 9$ –11 cilia.

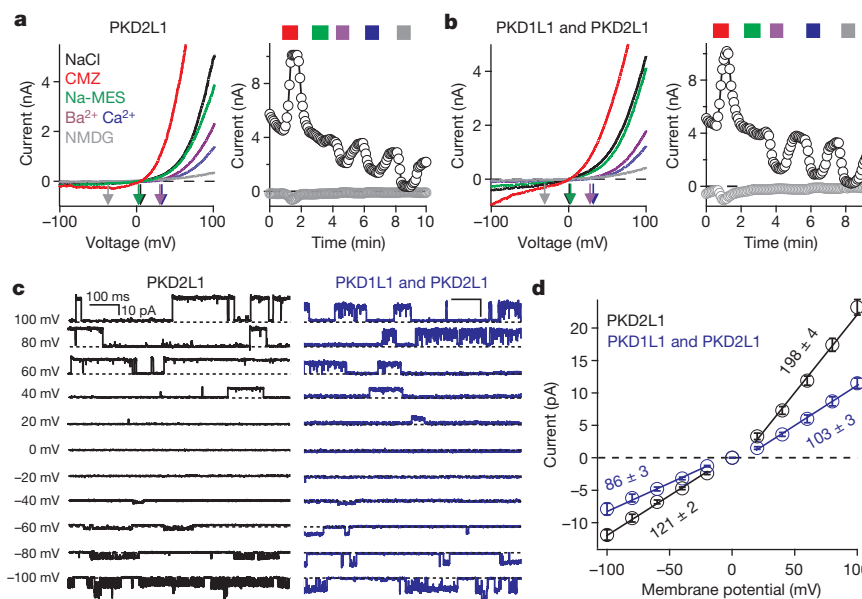


Figure 4 | Plasma membrane expressed PKD1L1–PKD2L1 channels match those of I_{cilia} . **a, b**, Whole-cell currents recorded from cells transfected with PKD2L1 alone (**a**) or PKD1L1 and PKD2L1 (**b**), where extracellular 1 μM calmidazolium (CMZ, red trace) was applied and Na⁺-based saline was exchanged by the cations indicated. **c**, Single-channel currents measured from

the cell membrane of cells transfected by PKD2L1 (black traces), or PKD1L1 and PKD2L1 (blue traces). **d**, Resulting average single-channel amplitudes plotted against voltage measured from transfected cells (\pm s.e.m., $n = 4$ –5 cells).

Immunoprecipitation of Flag- or haemagglutinin (HA)-tagged PKD1L1 and PKD2L1 demonstrated that PKD1L1 and PKD2L1 interact (Extended Data Fig. 4a). In whole-cell patch clamp of HEK293 cells transiently transfected with members of the PKD family, only PKD2L1 produced a measurable outward current in the plasma membrane (Extended Data Fig. 4b). Together, these data suggest that PKD2L1, but not PKD1L1, can form homomeric channels. On the basis of an alignment of the PKD2 family of proteins (Extended Data Fig. 4c), a cluster of conserved acidic residues in the putative selectivity filter¹⁷ was mutated. Neutralizing two of three extracellular glutamates (Glu 523 and Glu 525, but not Glu 530) to serine or alanine abolished the current (Extended Data Fig. 4b). The homomeric PKD2L1 channel produced an outwardly rectifying, cationic whole-cell current with a large moderately Ca²⁺-selective conductance (Fig. 4a and Extended Data Fig. 4d). Our results differ from the presumed PKD2L1 expressed in oocytes¹⁸: PKD2L1 currents in our HEK293T cells were outwardly rectifying and not activated by extracellular calcium. The unitary conductance (198 ± 3 pS) and rectification ratio of homomeric PKD2L1 is too large to underlie the primary ciliary conductance (compare Figs 1e and 4d). However, co-expression of PKD1L1 and PKD2L1 yielded a less rectifying whole-cell current with a single channel outward conductance of 103 ± 3 pS (Fig. 4b–d), consistent with that found in RPE primary cilia. In addition, heterologously expressed PKD2L1 and PKD1L1 channels were activated by calmodulin antagonists and blocked by Gd³⁺ and ruthenium red. Unlike the ciliary current, the heterologously expressed channels were not activated by extracellular uridine and purine analogues (UDP, ADP, ATP; data not shown), suggesting that the receptor/signal transduction pathway in cilia is absent in the whole-cell HEK293 conditions. These data establish that I_{cilia} is conducted by a heteromer of PKD1L1 and PKD2L1 subunits.

Primary cilia have been proposed to sense flow by mechanical activation of the putative PKD1–PKD2 channel^{16,19}. Using pressure clamp²⁰, we observed no difference in single-channel activity at pressures of 0–60 mm Hg (8 kPa) for PKD1L1–PKD2L1 in mRPE primary cilia or in heterologous expression (Extended Data Fig. 5). Although channel activity measurably increased at high pressure (80–100 mm Hg; 11–13 kPa), the PKD1L1–PKD2L1 channel lacks the sensitivity found in designated mechanosensitive channels^{21,22}.

Many members of the TRP superfamily are highly temperature sensitive^{7,23–25} and primary cilia are proposed to be temperature sensors via activation of a thermosensitive Ca²⁺ influx²⁶. We tested the effect of increasing temperature (22–37 °C) on primary cilia and PKD1L1–PKD2L1-transfected cells and observed an increase in the Gd³⁺-sensitive current (Extended Data Fig. 6a, b). When rapidly increasing the bath temperature (2.1 ± 0.2 °C s^{−1}), we observed biphasic current activation in the current from RPE primary cilia (from 24 to 32 °C, $Q_{10} = 6$) and heterologously expressed PKD1L1–PKD2L1 cilia (from 24 to 32 °C, $Q_{10} = 8$; Extended Data Fig. 6c). This sensitivity is moderate in comparison to highly temperature-sensitive channels, such as TRPV1–TRPV4 ($Q_{10} > 20$). In any case, temperature gradients are probably inconsequential over the length of cilia and cilia to cytoplasm.

The accompanying manuscript² demonstrated that cilia are a specialized calcium compartment containing the PKD1L1–PKD2L1 channel, which regulates ciliary SMO signalling and *Gli* transcription. Here we have demonstrated that PKD1L1 and PKD2L1 heteromultimerize to form a calcium-permeant ciliary channel which can be indirectly activated by purines. Future experiments will determine the purine receptor and downstream effectors activating I_{cilia} . Because the PKD1L1–PKD2L1 complex is calcium-permeant, but is also inactivated at high internal Ca²⁺ concentration, we propose that the channel regulates the high resting ciliary calcium level². We propose that these signals modify regulation of smoothed target genes, in particular *Gli* transcription of genes regulating cell division and growth.

METHODS SUMMARY

Cilia were recorded via borosilicate glass electrodes polished to resistances of 13–19 M Ω . Holding potentials were -60 mV. Unless stated otherwise, pipette solutions contained (in mM): 100 CsMES, 35 NaCl, 10 HEPES, 5 Cs-BAPTA, 2 MgCl₂, 100 nM free [Ca²⁺], pH 7.4. Standard bath solution (SBS) contained 150 NaCl, 10 HEPES, 1.8 CaCl₂, 1 MgCl₂, pH 7.4. For 'on cilia' recordings, the standard bath solution was: 150 KCl, 10 HEPES, 1.8 CaCl₂, 1 MgCl₂, pH 7.4 to clamp the resting membrane potential. For on-cilia recordings, the intracellular pipette solution was SBS. Perforated patches were obtained using 0.4 mg amphotericin B per 1 ml of standard intracellular saline. siRNA knockdown efficiency was monitored for every set of experiments with Silencer Negative Control 1 siRNA (Life Technologies) for controls. Sequences for gene-specific primers are listed in Extended Data Fig. 3a.

Additional experimental procedures are described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 March; accepted 8 November 2013.

- Bornens, M. The centrosome in cells and organisms. *Science* **335**, 422–426 (2012).
- Delling, M., DeCaen, P. G., Doerner, J. F., Febvay, S. & Clapham, D. E. Primary cilia are specialized calcium signalling organelles. *Nature* <http://dx.doi.org/10.1038/nature12833> (this issue).
- Corbit, K. C. *et al.* Vertebrate Smoothed functions at the primary cilium. *Nature* **437**, 1018–1021 (2005).
- Hardie, R. C. & Minke, B. The *trp* gene is essential for a light-activated Ca^{2+} channel in *Drosophila* photoreceptors. *Neuron* **8**, 643–651 (1992).
- Shin, J. B. *et al.* *Xenopus* TRPN1 (NOMPC) localizes to microtubule-based cilia in epithelial cells, including inner-ear hair cells. *Proc. Natl Acad. Sci. USA* **102**, 12572–12577 (2005).
- Stortkuhl, K. F., Hovemann, B. T. & Carlson, J. R. Olfactory adaptation depends on the Trp Ca^{2+} channel in *Drosophila*. *J. Neurosci.* **19**, 4839–4846 (1999).
- Story, G. M. *et al.* ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112**, 819–829 (2003).
- Kleene, N. K. & Kleene, S. J. A method for measuring electrical signals in a primary cilium. *Cilia* **1**, 1–17 (2012).
- Arnaout, M. A. Molecular genetics and pathogenesis of autosomal dominant polycystic kidney disease. *Annu. Rev. Med.* **52**, 93–123 (2001).
- Celic, A., Petri, E. T., Demeler, B., Ehrlich, B. E. & Boggon, T. J. Domain mapping of the polycystin-2 C-terminal tail using *de novo* molecular modeling and biophysical analysis. *J. Biol. Chem.* **283**, 28305–28312 (2008).
- Yu, Y. *et al.* Structural and molecular basis of the assembly of the TRPP2/PKD1 complex. *Proc. Natl Acad. Sci. USA* **106**, 11558–11563 (2009).
- Zhu, J. *et al.* Structural model of the TRPP2/PKD1 C-terminal coiled-coil complex produced by a combined computational and experimental approach. *Proc. Natl Acad. Sci. USA* **108**, 10133–10138 (2011).
- Gherman, A., Davis, E. E. & Katsanis, N. The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia. *Nature Genet.* **38**, 961–962 (2006).
- Andrade, Y. N. *et al.* TRPV4 channel is involved in the coupling of fluid viscosity changes to epithelial ciliary activity. *J. Cell Biol.* **168**, 869–874 (2005).
- Raychowdhury, M. K. *et al.* Characterization of single channel currents from primary cilia of renal epithelial cells. *J. Biol. Chem.* **280**, 34718–34722 (2005).
- Yoshida, S. *et al.* Cilia at the node of mouse embryos sense fluid flow for left-right determination via Pkd2. *Science* **338**, 226–231 (2012).
- Hille, B. *Ion Channels of Excitable Membranes* 3rd Edn (Sinauer Associates, 2001).
- Chen, X. Z. *et al.* Polycystin-L is a calcium-regulated cation channel permeable to calcium ions. *Nature* **401**, 383–386 (1999).
- Praetorius, H. A. & Spring, K. R. A physiological view of the primary cilium. *Annu. Rev. Physiol.* **67**, 515–529 (2005).
- Besch, S. R., Suchyna, T. & Sachs, F. High-speed pressure clamp. *Pflügers Arch.* **445**, 161–166 (2002).
- Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, 176–181 (2012).
- Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R. & Kung, C. A large-conductance mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature* **368**, 265–268 (1994).
- Xu, H. *et al.* TRPV3 is a calcium-permeable temperature-sensitive cation channel. *Nature* **418**, 181–186 (2002).
- Benham, C. D., Gunthorpe, M. J. & Davis, J. B. TRPV channels as temperature sensors. *Cell Calcium* **33**, 479–487 (2003).
- Voets, T. *et al.* The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* **430**, 748–754 (2004).
- Kottgen, M. *et al.* TRPP2 and TRPV4 form a polymodal sensory channel complex. *J. Cell Biol.* **182**, 437–447 (2008).
- Cang, C. *et al.* mTOR regulates lysosomal ATP-sensitive two-pore Na^{+} channels to adapt to metabolic state. *Cell* **152**, 778–790 (2013).
- Fieni, F., Lee, S. B., Jan, Y. N. & Kirichok, Y. Activity of the mitochondrial calcium uniporter varies greatly between tissues. *Nature Commun.* **3**, 1317 (2012).
- Shenkel, S. & Sigworth, F. J. Patch recordings from the electrocytes of *Electrophorus electricus*. Na currents and PNa/PK variability. *J. Gen. Physiol.* **97**, 1013–1041 (1991).
- Brisson, A. & Unwin, P. N. Quaternary structure of the acetylcholine receptor. *Nature* **315**, 474–477 (1985).

Supplementary Information is available in the online version of the paper.

Acknowledgements P.G.D. was supported by NIH T32 HL007572. Animal work was, in part, supported by NIH P30 HD18655 to the IDRC of Boston Children's Hospital. We thank B. Navarro, N. Blair, J. Doerner, S. Febvay, and the members of the Clapham laboratory for advice and assistance.

Author Contributions All authors designed or conducted experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.C. (dclapham@enders.tch.harvard.edu).

CAREERS

COLUMN Negative results can be interesting — but they still hurt **p.321**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com



COMPUTING

Out of the hood

Biologists frustrated with wet-lab work can find rewards in a move to computational research.

BY ROBERTA KWOK

During her master's programme in genetics from 2005 to 2008, Sarah Hird dreaded going into the lab. She was studying subspecies of red-tailed chipmunks and had become discouraged and frustrated by the uncertainties of molecular-biology experiments. She spent six weeks trying to amplify repetitive sequences in chipmunk DNA as part of an experiment to identify genetic differences between populations — but to no avail. Hird tried replacing

reagents, switching to a different machine for running the polymerase chain reaction and decontaminating the sample-preparation area. Nothing worked. And then, for reasons that she never quite deciphered, the technique suddenly started working again.

By the end of her master's, Hird had come to dislike working in a wet lab, and she decided not to apply for PhD programmes.

About six months after finishing her master's degree, while working as a part-time technician at Louisiana State University in Baton Rouge, she discovered a better

direction. The lab's principal investigator had suggested that she learn a computer-programming language so that she could help with a simulation project. Hird, who had never programmed before, taught herself the language using a book and online tutorials, and quickly became engrossed. "Once I started, it was like an addiction," she says. She enjoyed developing algorithms, and she found the software-debugging process less frustrating than troubleshooting wet-lab problems. The work felt more under her control.

Hird decided to stay in science, but to focus on dry-lab research — which generally entails writing software and developing models to analyse or explain data. After gaining admission to Louisiana State University's PhD programme in evolutionary biology, Hird developed software programs to manipulate sequence data for her dissertation. To boost her computer-programming skills, she took an undergraduate computer-science course and enrolled in a summer software-development programme. This year, she started a postdoc on computational analysis of bacterial sequences at the University of California, Davis.

Hird's success demonstrates that a wet-lab biologist can make a smooth transition to the dry lab, even without extensive formal computer-science education. Resources for learning computational biology and bioinformatics abound: there are training websites, online forums and workshops. And motivated scientists with a knack for quantitative thinking and problem solving can gain the skills necessary to make the change during a PhD programme or postdoc. "A lot of people get scared about even trying," says Casey Bergman, a computational biologist at the University of Manchester, UK. "It's a much easier transition than most people think."

Dry-lab researchers enjoy lower overhead costs than wet-lab scientists and have the flexibility to start new projects quickly. They can also more easily continue to run experiments when they become principal investigators, whereas scientists in wet labs tend to move entirely into managerial roles. Computational biologists face their own problems — some feel that they are taken less seriously than wet-lab scientists, for example — but those who set their own research directions and establish productive collaborations can thrive. For a researcher who wants to conduct experiments with quick turnaround times, and would like the ability to work on a wide array of topics, the flexibility to do research from home on a ►

► computer and the option to move into jobs in software development or data analysis, the dry lab can be a promising career option.

LEARNING TO CODE

A wet-lab scientist without a programming background can take courses in computational biology or software programming at university. But for those who want to try out the dry lab without committing a lot of time, there are several do-it-yourself options. The website Galaxy (galaxyproject.org), for example, offers a user-friendly way to run computational-biology analyses. Users can upload DNA- or RNA-sequence data and use a point-and-click interface to run software, such as a tool to find regions of DNA in which promoters overlap binding sites for transcription factors. A molecular biologist might analyse genetic data from their wet-lab experiments or the sample data sets available on the website.

The site, developed by researchers at Pennsylvania State University in University Park, and Emory University in Atlanta, Georgia, also provides an interactive tour that demonstrates how to carry out analyses. Galaxy is good for beginners because it allows them to try bioinformatics tools without writing software programs, says Bergman.

To start programming, Hird suggests reading a book for beginners and doing the exercises. Websites such as Codecademy

(codecademy.com) offer online tutorials. Aspirants might start with an easy language such as Python instead of more difficult languages such as Java or C++, says Pavel Tomancak, a developmental biologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany.

He notes that newcomers may stay more motivated if they can apply computational skills to real scientific problems rather than to the 'toy' exercises in a computer-science class. For example, a researcher who works with many image files could write a program to automatically perform processing steps, such as contrast enhancement, on thousands of images.

Biologists can practise their programming skills on websites such as Rosalind (rosalind.info), developed by researchers at the University of California, San Diego, and at St Petersburg Academic University in Russia. The site presents progressively harder bioinformatics problems and asks the user to write programs to solve them. An early problem involves converting a DNA string into its transcribed RNA strand; a more advanced exercise requires the user to find ways in which RNA molecules can bind to each other.

Users download sample input files that contain the data to be processed, run their programs on those data and copy the output to the site to confirm whether their answers



"Once I started, it was like an addiction."

Sarah Hird

were correct. Once users have completed an exercise successfully, they can view solutions that other users have written. Those who can make a bigger time commitment can pursue more-formal training options (see 'Classes in computing').

Beginners can also ask questions on online bioinformatics, sequencing

and programming forums such as Biostar, SEQanswers and Stack Overflow. "If you're feeling a bit stuck, there's no shortage of help," says Neil Saunders, a statistical bioinformatician at the Commonwealth Scientific and Industrial Research Organisation in Sydney, Australia. To avoid receiving snide responses, newcomers should make their questions specific and describe their efforts to solve problems instead of simply asking others to write their code.

TIME FOR A CHANGE

Scientists seeking a transition to the dry lab can build their computational skills during their wet-lab PhD programme. They might consider using computational tools in their dissertation work, as well as seeking out a supportive mentor or committee member, says Hird. To gain an adviser's support, a doctoral student could explain how these skills would enable better research: for example, a graduate student in developmental biology could use programming skills to perform complex three-dimensional data visualizations. Researchers could also collaborate with dry labs to analyse their data, suggests Karen Cranston, a computational evolutionary biologist at the National Evolutionary Synthesis Center in Durham, North Carolina.

PhD graduates can land dry-lab postdocs without extensive programming experience. Bergman finds that applicants with the perfect blend of skills in biology and computer science are rare, so he is willing to hire and train biologists who do not know programming as long as they have strong quantitative skills. Nevertheless, applicants can improve their chances by using Rosalind, learning a statistical software package such as R or completing a computational project through a programme such as Google Summer of Code.

Computational biologists tend to enjoy more flexibility than their wet-lab counterparts. Researchers can often start a new project by downloading data and writing a program, instead of having to order reagents or spend months optimizing protocols. Costs are generally lower because dry labs do not require reagents or staff to maintain cell lines, and

FORMAL TRAINING

Classes in computing

Researchers seeking formal training in computational biology or bioinformatics can enrol in classes, workshops and mentoring programmes. Sarah Hird, a postdoctoral researcher in evolutionary biology at the University of California, Davis, gained computational skills during her PhD programme by enrolling in 2011 in Google Summer of Code. The programme, run by Google in Mountain View, California, funds project-proposal-based software-development projects by undergraduates and graduate students, and links participants with mentoring organizations. Hird learned the programming language Python and developed software to display features of DNA-sequence data in a tabular format, receiving guidance from two experienced programmers through e-mails, Skype calls and comments on her blog posts about the project.

Last month, researchers at the University of California, San Diego, began teaching a free online course on bioinformatics algorithms through the education website Coursera (www.coursera.org). They hope to repeat the course next year. Many

institutions run computational-biology workshops, including Cold Spring Harbor Laboratory in New York, the National Evolutionary Synthesis Center (NESCent) Academy in Durham, North Carolina, and the European Molecular Biology Laboratory in Heidelberg, Germany. Karen Cranston, a computational evolutionary biologist at NESCent, suggests signing up for mailing lists to find out about upcoming workshops.

Software Carpentry, an international volunteer organization, offers software-development workshops for scientists worldwide. The two-day boot-camps help researchers learn efficient programming, such as managing software versions and setting up automated bug testing. Researchers wanting their institution to host a workshop should contact Software Carpentry, says Titus Brown, an advisory-board member and bioinformatician at Michigan State University in East Lansing. The organization can provide a 'pitch' presentation that contains information about benefits and costs, which the researchers can show to administrators to convince them to fund instructors' travel expenses. **R.K.**

results come back quickly, making it less risky to take research in a new direction. When a computer is the only equipment needed, one does not always have to work in one's office. Bergman often writes software or runs programs at home in the evenings. "That's what sustains my interest in the science," he says.

Dry-lab researchers can apply their skills to many problems. "It really expands what you can do," says Saunders. "If your speciality is analysing data, then the nature of the data is sort of secondary." Saunders has analysed the genes of Antarctic microbes, modelled protein structures to study how enzymes find the right substrates and searched for biomarkers to improve early detection of colorectal cancer. And dry-lab scientists have plenty of options outside academia: career opportunities are opening up at institutions such as museums, which need staff with computational skills to help to organize and share their data, such as biodiversity information, and at DNA-sequencing facilities. Bergman and Saunders know of dry-lab researchers who have gone on to work at genomics and online education start-ups in California's San Francisco Bay Area, or to develop mobile-phone apps or land data-analysis jobs at consulting companies.

Dry-lab scientists must ensure, however, that wet-lab collaborators do not view them only as technicians — they must make it clear that they are "actually a research scientist", says Saunders. "You're not just the computer guy with the magical program." Even when one is the principal investigator of a lab, it is easy to be perceived as just "providing a service", he says. So it is important that, when starting a collaboration, computational researchers discuss whether they will help to set the project's scientific direction. Saunders adds that it is fine to participate in some projects that are steered by wet-lab scientists, but dry-lab scientists should develop their own projects as well, such as developing computational tools.

Researchers who make the transition to the dry lab need not worry that their wet-lab experience is wasted. Dry-lab scientists still need to think like biologists and consider the complexities of the system being studied. Programmers tend to oversimplify complex problems, but biologists know that there are often exceptions to the rules, says Cranston. And some researchers combine wet-lab and dry-lab work rather than giving up the former entirely. "I don't really like to look at it like a switch. I look at it more like adding more tools to your toolkit," says Bergman. "It's not an either/or." ■

Roberta Kwok is a freelance science writer in Seattle, Washington.

COLUMN

A faulty hypothesis

Carolyn Beans says that some of the most interesting results are negative ones — but it still hurts to be wrong.



I hold a strip of tinfoil above a bowl of water and ask the four children seated around me to make a hypothesis. "What do you think? Will it sink or will it float?" They have all sorts of ideas. "Float! It's too light to sink." "Sink! It's made of metal!" "Float! It's flat!"

I ask the students to circle their guesses next to the picture of the tinfoil in the hypothesis column of their worksheets. Then I drop the foil. The two students who circled 'float' begin flailing their arms in a wild dance of joy. The two who circled 'sink' erase their choice.

"Don't erase!" I beg them. "Just circle 'float' in the results column. Some of the most interesting scientific discoveries come when a hypothesis is wrong." They do not buy it. As we test more objects, the children continue to modify their hypotheses to match the results. One refuses to make a hypothesis until the results are known. Apparently, children do not like to be wrong, even in the name of science.

I thought of this elementary-school visit as I sat in my office last autumn, looking over the results of my summer research. I had tested whether three different plant species could reduce seed production in a jewelweed native to the United States, *Impatiens capensis*, by competing for pollinators.

One of the contenders was an introduced jewelweed from India, *Impatiens glandulifera*, with bright-magenta, nectar-rich flowers. I thought that surely this species would win over the local pollinators and reduce seed production in the native plant.

But it turns out that none of the competitors, either individually or as a group, had any influence on seed production in the native

jewelweed. When I first saw these results, I thought how nice it would be to erase the past four months of work and start again.

As a graduate student, I find it hard not to feel like a failure when a hypothesis is incorrect. It was especially disheartening when I had committed an entire summer to an experiment, and there was no hope of restarting until the next growing season. Throughout the autumn and winter, I pondered. Was my sample size too small? Were my plots too close together? Should I have run the experiment for longer?

I came to the conclusion that my hypothesis was just wrong. Although this realization did not leave me dancing for joy like the children who circled 'float' (and I certainly would have danced had my hypothesis been correct), it did allow me to start identifying the positive aspects of my results.

First, I feel grateful that the experiment happened at all. Fieldwork is unpredictable and sample sizes are often demolished when study sites succumb to disease, unusual weather patterns or a stray lawnmower. Getting 700 potted plants of 4 species to flower all at the same time was no trivial task. For my success I thank good field assistants, decent weather and an influx of ladybirds that appeared just when it seemed that aphids would eat every last plant.

Second, I appreciate that my experiment was designed well enough to disprove my hypothesis definitively. I did not get what I expected, but I discovered something real. Now I can develop new experiments instead of wasting time redoing the same one.

Finally, I feel hopeful that the negative results mean that the story will turn out to be more nuanced and interesting than I initially expected. For example, this and other experiments suggest that even without influencing seed production, competitors may alter natural selection on the native jewelweed by causing pollinators to favour different-sized flowers.

As for the kids, they still want to be right. But what they want most is to do another experiment, and then another. They never tire of dropping objects into the water and thinking up reasons that one will float and another will not. They may not own up to their incorrect hypotheses. But they are learning that there is more to science than confirming one's expectations. ■

Carolyn Beans is a biology graduate student at the University of Virginia in Charlottesville.

MELANIE TAYLOR/SHUTTERSTOCK

SIBYL

Life can be a drag.

BY DEBORAH WALKER

The ghost of my future smells of ash. “I thought that you were going to stop smoking,” I say.

“It’s been a tough year.” She rummages inside her bag and produces a packet of Marlboro Lights. “Life doesn’t always go according to plan, does it, Sibyl?” She lights a cigarette and blows the smoke towards me. Ghost smoke, a multiplication of the insubstantial.

“I think I’ll join you.” I take a cigarette from my own packet while taking a critical look at my future self. She looks much older than she looked a year ago. She’s not doing herself any favours by not wearing make-up. Her hair looks dry and brittle, and the roots need doing. “I see that you haven’t lost any weight.”

She shrugs. “Dieting’s a waste of time. I’m nearly forty. I am what I am.”

She’s in one of those moods. “So, what’s new?” I ask.

“Not much.”

I sigh. “That’s not very helpful. This rite is not without sacrifice, you know.” I point to the iron knife balancing on top of the dish of blood water.

“Don’t I know it?” She rolls up her sleeve and shows me her right arm. She is seven years older than me, seven more scars. This is how it works, once a year I can see seven years into the future.

“Shall we do the diary?” I ask.

“Ah, yes, the diary.” She takes the leather diary out of her bag. I’d bought it in Venice, on my honeymoon. I’m supposed to write in it every day: the diary of my life.

The ghost flicks through the pages. “The trouble with this diary is that it gets a little sketchy in places. You’re drinking a lot at the moment, aren’t you?”

I shrug. I like a glass of wine or two in the evening. It takes the edge off. Who is she to judge me? “Shall we get on with the markets?”

“Sure.” My future self recites share prices while I take notes. I play the market. Although playing implies that I’ve a possibility of losing. That’s not the case, not with the information I’m receiving. I’m the ultimate insider dealer.

When she’s finished, she says: “All right then, I’ll be off.”

“Don’t go yet.”

“What is it?” she asks, impatiently.

“You don’t look great.”

“Thanks a lot.”

“I mean, what’s

happened to you in



the last year?” I feel sorry for her, but more importantly I feel anxious. I need to know.

“It’s best not to talk about personal stuff, Sibyl, you know that.”

“How’s Alex?”

“Are you sure you want to know?”

“It is Alex, isn’t it? What’s happened? He’s not . . . dead, is he?”

She lights another cigarette. I do the same.

“Alex left me.”

“But last year you seemed so happy.”

“Ignorance is bliss. He’s been having an affair for the last three years. Alice gave him an ultimatum, and I lost out.”

“Alice? My best friend Alice?”

“That’s right. He’s taking me through the courts now, trying to get his ‘fair share.’”

“I don’t believe it.”

“Would I lie to you? Would I lie to myself?”

She looks at me. “What are you going to do, now that you know?”

I walk to the fridge and pour myself a glass of cold, crisp chardonnay. I drain the glass.

She watches me with a half-smile. I refill the glass. “You shouldn’t have told me.”

“At least I gave you a warning. That’s more than I got.”

“She didn’t tell you?” Timelines are divergent. Each future me is slightly different.

“No. She didn’t. But I thought you’d want to know. That’s our trouble, we always want to know.” She blows a plume of ghost smoke towards me. “You could divorce him.”

“You had nine good years of marriage.”

“No, I didn’t. For three of those years Alex was having an affair.”

She lets her cigarette fall to the ground. “What are you going to do, Sibyl?” She has a hungry look on her face. She wants me to say that I’m going to divorce Alex, before he’s had a chance to betray me. When did I get so bitter?

“I don’t know what I’m going to do.”

“It’s your decision,” she says. “It won’t change anything for me. I’ll just carry on in this timeline where he betrayed me. You can’t change the past, only the future.”

“And you?” I say. “Are you going to look ahead, this year?”

“I always do, don’t I?” She rubs her arm. “Find out how I can improve my perfect life.”

“You don’t need to. You must have plenty of money stashed away.”

“No. I don’t need to look into the future. But then again, neither do you.”

“It’s a hard habit to break.”

She nods. I see the shadow in her eyes. I know her fear. The same fear that shrouds me every time I start the ritual. There will come a day when I reach into the future and my future self will be dead. What will I see on that night? Will I see nothingness, or something worse, something unbearably worse?

“I’m young,” she says. “I’m only thirty-eight. It will be okay to look.”

“Yes. It’ll be okay. Thanks for your help.”

“It’s nothing. Be well, Sibyl. Be happy.”

With a word I end the ritual, and my future self dissipates.

I tidy up, throwing the blood water down the sink and washing the bowl. Alex would be home soon. Could I change, make our marriage stronger? Did I want to?

A key rattles in the lock. Alex is home.

What could I say to him?

Divination is a drug.

I reach for the packet of cigarettes. Tomorrow, I’ll quit. ■

Find Deborah in the British Museum, trawling the past for future inspiration.

JACEY

➔ NATURE.COM

Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm